Phenix workshop at UT Southwestern Medical Center,
May 20, 2026

Molecular Replacement

Dorothee Liebschner
Lawrence Berkeley Laboratory

Crystallography: Phase problem

$$\rho(\boldsymbol{r}) = \frac{1}{V_{cell}} \sum_{hkl} |F(\mathrm{hkl})| e^{i\alpha} e^{-2\pi i \boldsymbol{hr}}$$

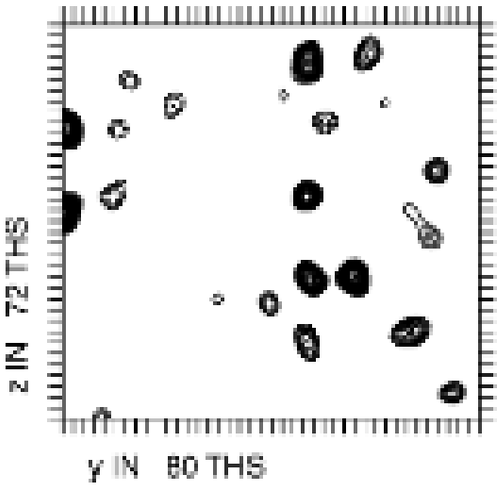
amplitude

phase

Measured in diffraction experiment

Not measured

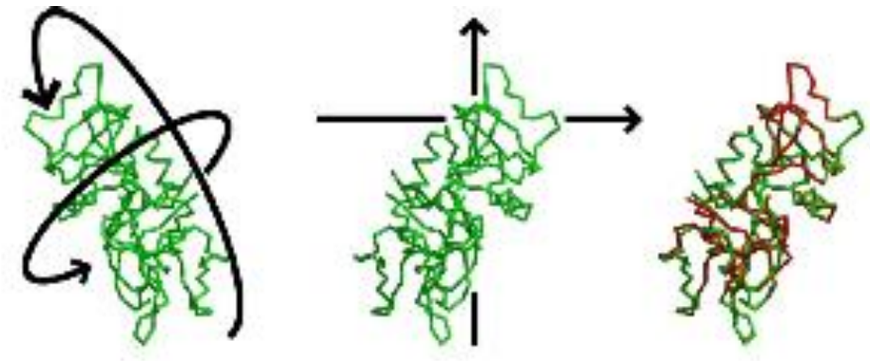
How to recover phases in Crystallography



Experimentally

Exploit the properties of a few special atoms:

- Anomalous scattering
- A large number of electrons



Computationally

- *Molecular Replacement (MR)*
  A previously known structure provides initial phase estimates for a new structure

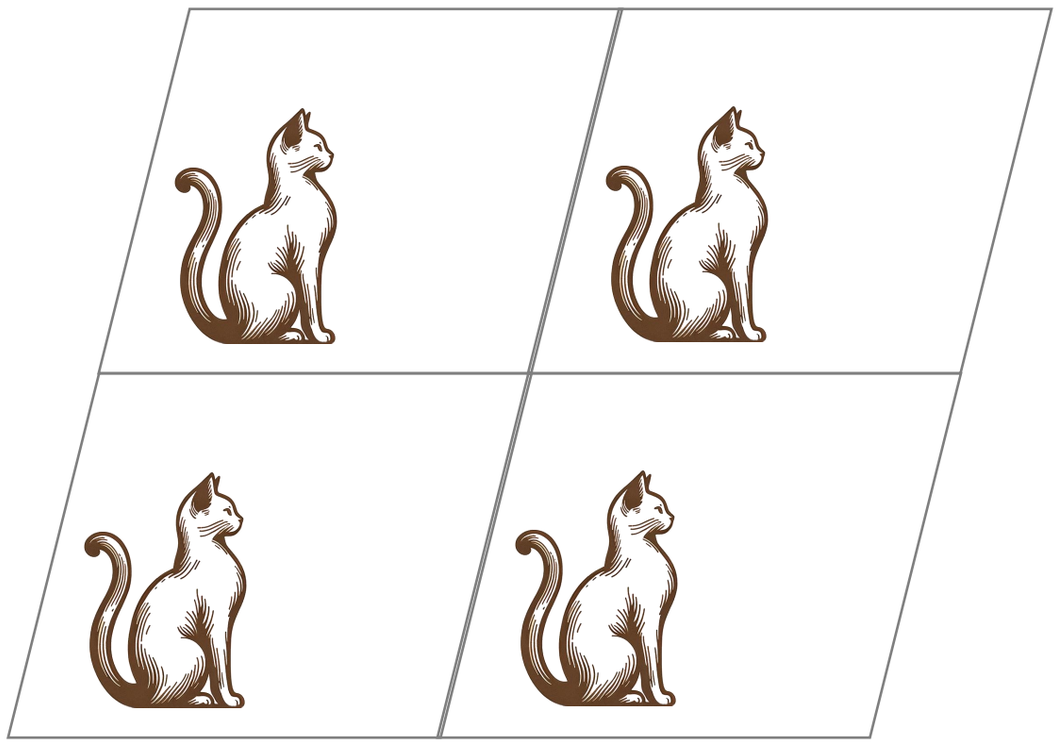
- *Direct Methods*
  Phase relationships can be formulated by assuming the positivity and atomicity of the electron density

Molecular Replacement (MR)

MR = solve the unknown crystal structure of a molecule using a related known molecular model.

Known model

Crystal of unknown structure

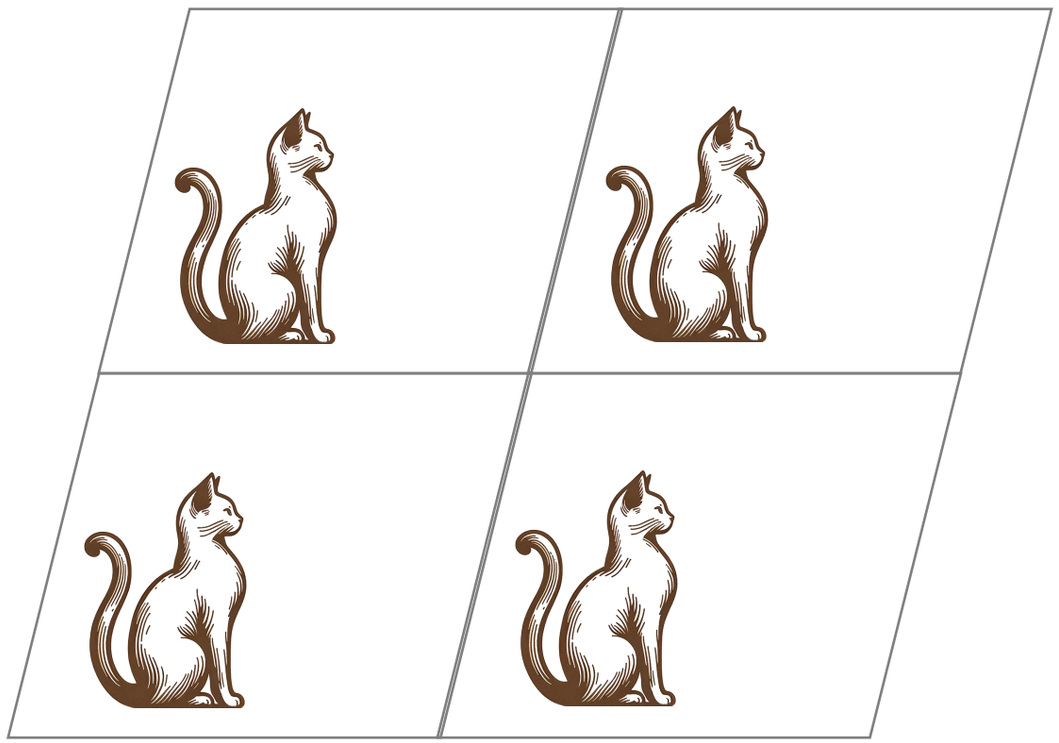


Molecular Replacement (MR)

MR = solve the unknown crystal structure of a molecule using a related known molecular model.

Known model

Crystal of unknown structure

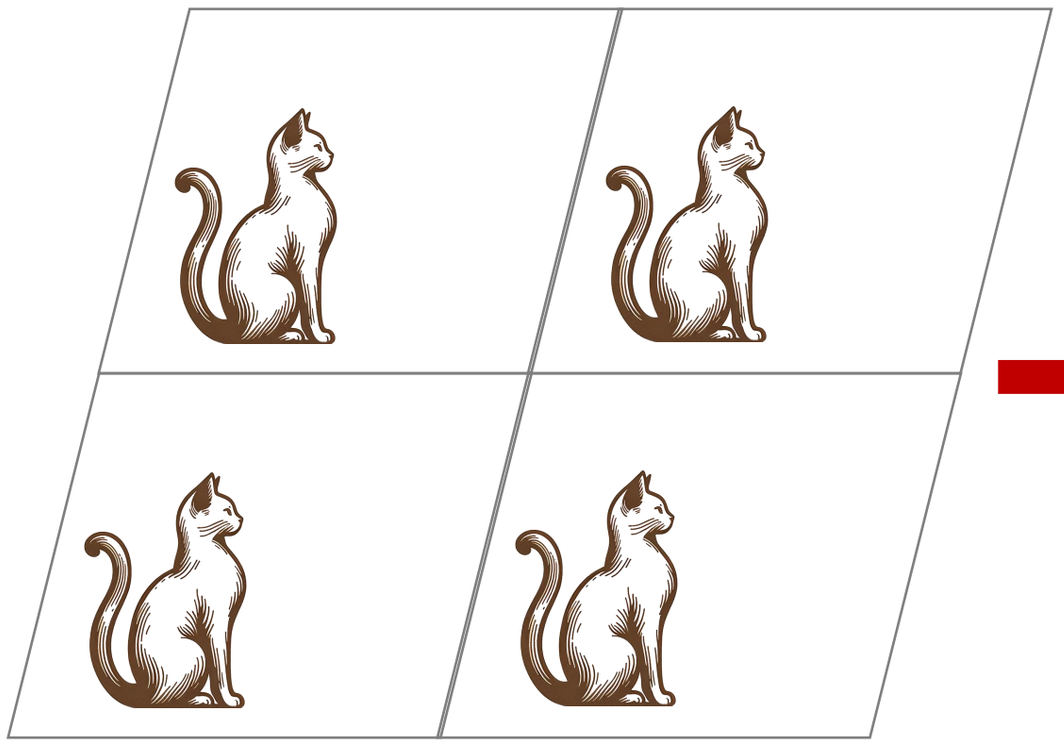


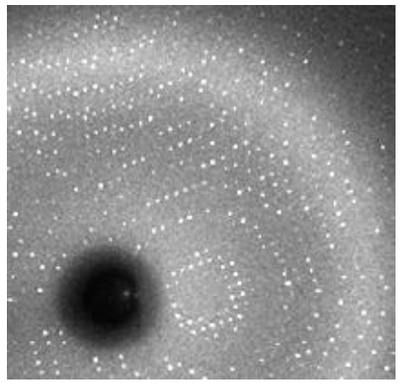
Known model provides initial estimates of the phases of the unknown structure.

Molecular Replacement (MR)

Known model

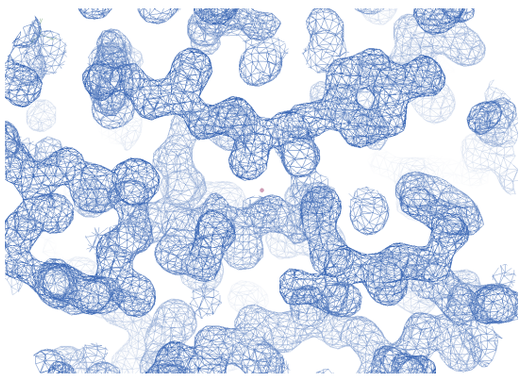
Crystal of unknown structure





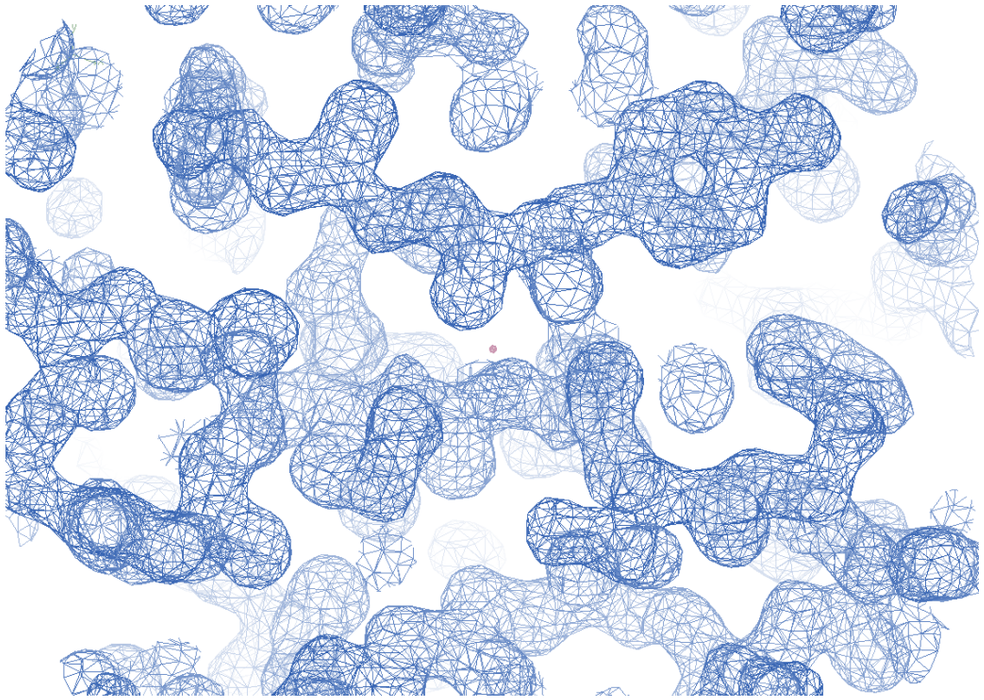
Intensities (hkl)

$|F|\ e^{i\phi}$



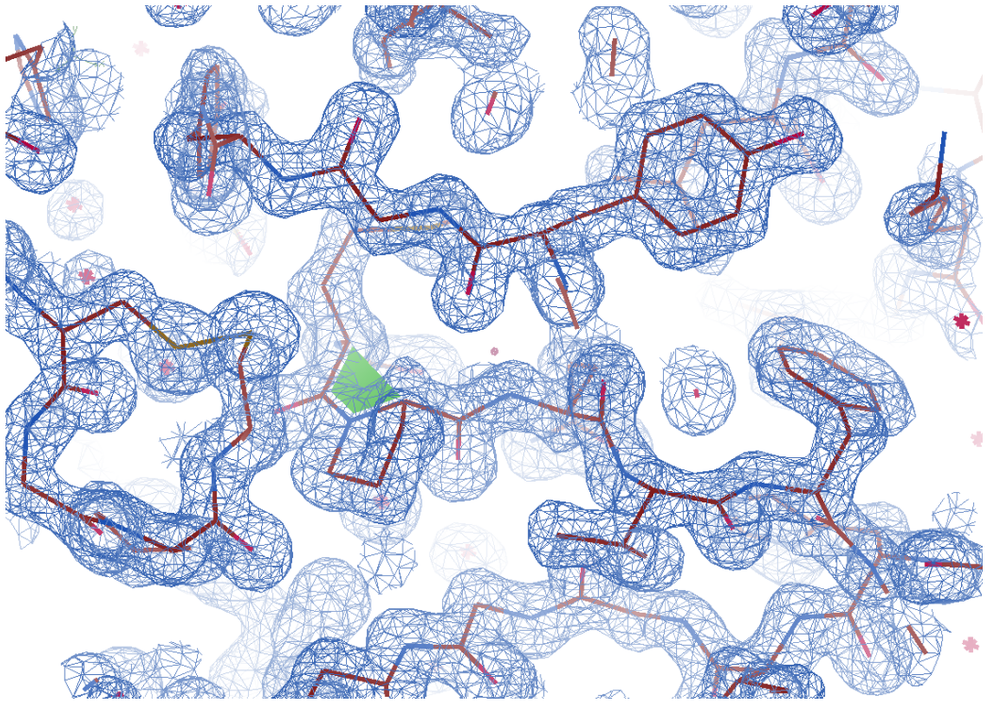
Density map

Molecular Replacement (MR)



If we know the density…

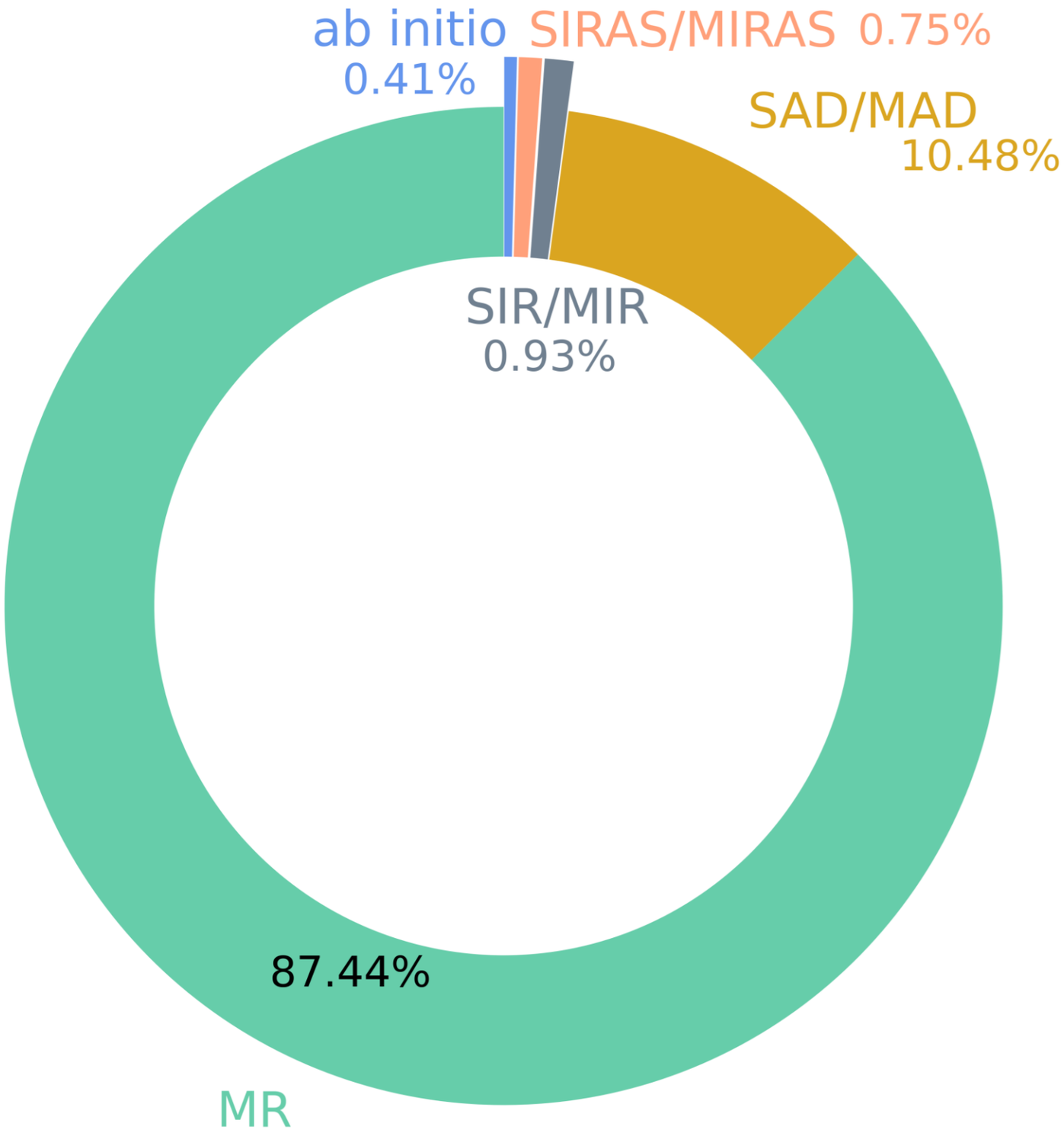
Molecular Replacement (MR)



If we know the density…

… then we can
determine the structure

Phasing methods in the PDB



MR method:

- Fast
- Cheap
- Highly automated

Known structure:

- Large number of deposited structures (PDB: 234k)
- Predicted models

Note: Not all models in the PDB have info, and even if present, it may not represent major phasing method.

MR will solve most structures

Try to solve recent SAD structures with MR and AF model.

STRUCTURAL BIOLOGY

ISSN: 2059-7983

Volume 79 | **Part 3** | March 2023 | Pages 234-244
https://doi.org/10.1107/S205979832300102X
OPEN ACCESS
Cited by 50
Part of a special issue

Accelerating crystal structure determination with iterative *AlphaFold* prediction

Thomas C. Terwilliger,[a,b*] Pavel V. Afonine,[c] Dorothee Liebschner,[c] Tristan I. Croll,[d] Airlie J. McCoy,[d] Robert D. Oeffner,[d] Christopher J. Williams,[e] Billy K. Poon,[c] Jane S. Richardson,[e] Randy J. Read[d] and Paul D. Adams[c,f]

[a]New Mexico Consortium, Los Alamos, NM 87544, USA, [b]Los Alamos National Laboratory, Los Alamos, NM 87545, USA, [c]
[d]Depa
United
Univer
*Corre

90% of 215 data sets solved with MR

STRUCTURAL BIOLOGY

ISSN: 2059-7983

Volume 80 | **Part 11** | November 2024 | Pages 766-779
https://doi.org/10.1107/S2059798324009380
OPEN ACCESS
Cited by 2
Part of a special issue

The success rate of processed predicted models in molecular replacement: implications for experimental phasing in the *AlphaFold* era

Ronan M. Keegan,[a,b‡] Adam J. Simpkin[a‡] and Daniel J. Rigden[a*]

[a]Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool L69 7ZB, United Kingdom, and [b]UKRI–STFC, Rutherford Appleton Laboratory, Research Complex at Harwell, Didcot OX11 0FA, United Kingdom
*Correspondence e-mail: drigden@liv.ac.uk

97% of 406 data sets solved with MR

(6% are “harder” cases)

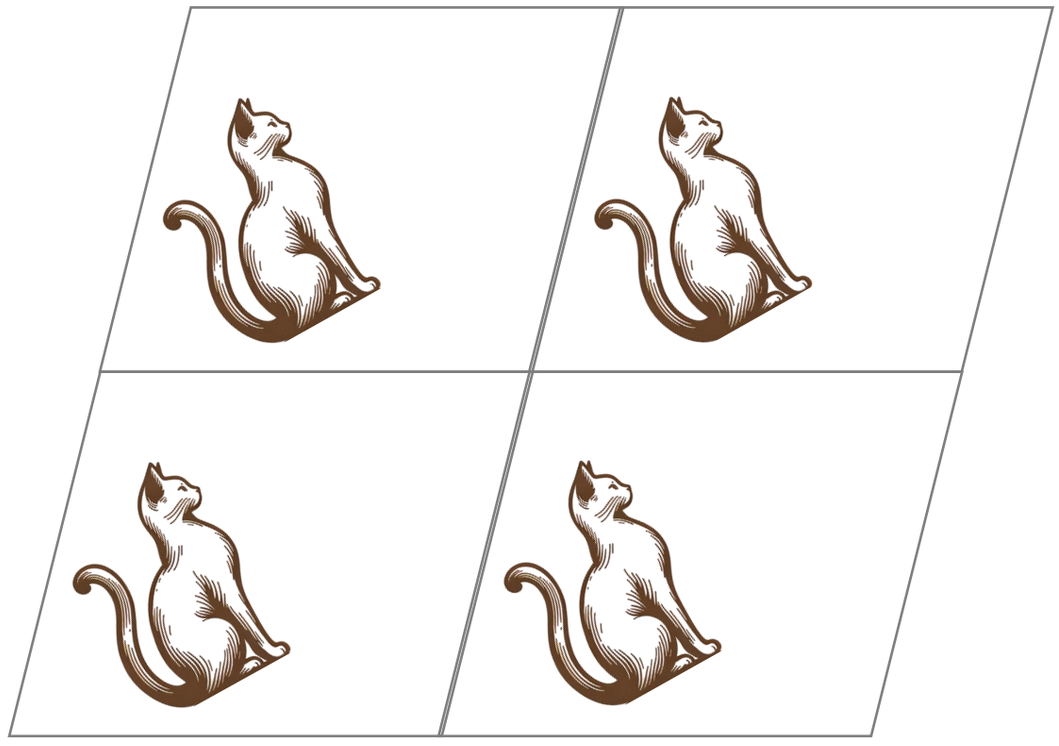
Terwilliger, T. C. *et al*. *Acta Crystallogr. D Struct. Biol.* **79**, 234–244 (2023).
https://doi.org/10.1107/S205979832300102X
Keegan, R. M., Simpkin, A. J. & Rigden, D. J. *Acta Crystallogr. D Struct. Biol.* **80**, 766–779 (2024).
https://doi.org/10.1107/S2059798324009380

Molecular replacement: Approach

Try to match the known model with the unknown structure.

Search model

Crystal of unknown structure

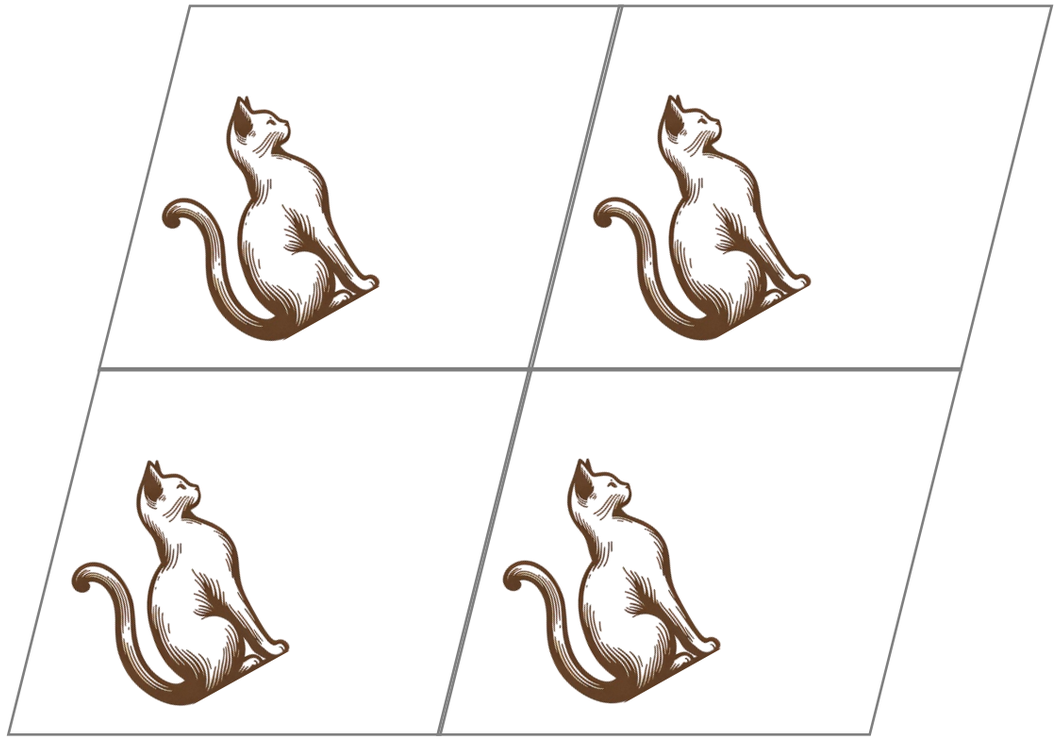


Find the rotation and translation of the search model so that it matches the unknown structure.

The search model

- Finding a suitable search model is critical step in MR.
- Should provide a high proportion of the scattering from the target structure with high accuracy (low r.m.s.d.).

Crystal of unknown structure

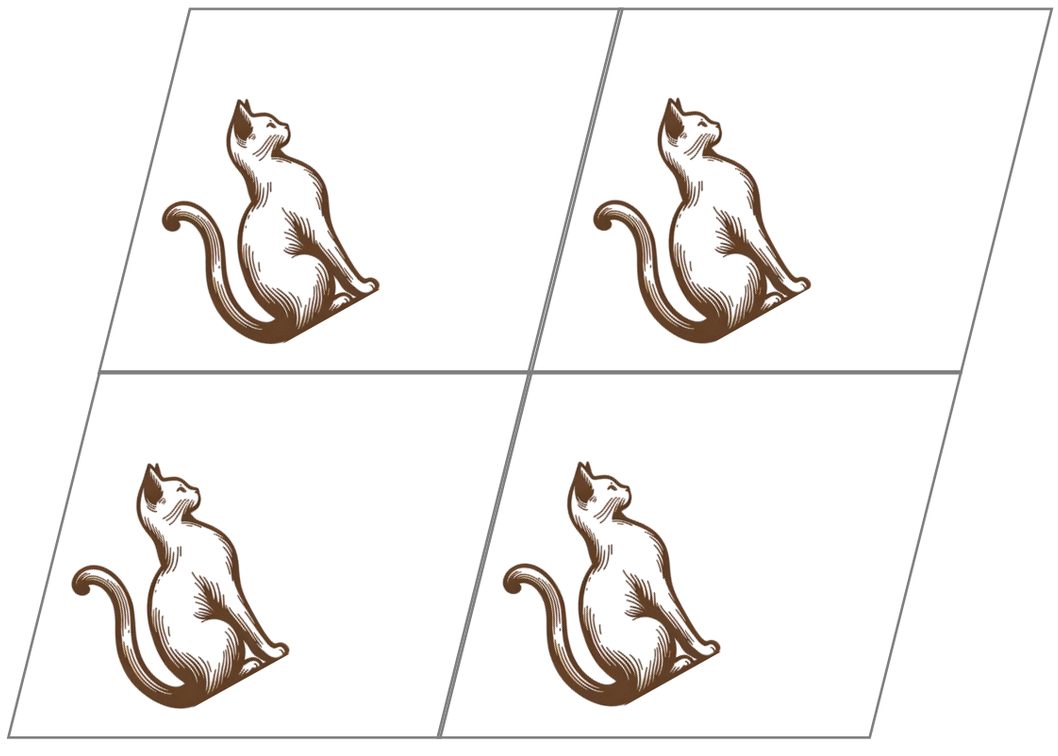


The search model

- Finding a suitable search model is critical step in MR.
- Should provide a high proportion of the scattering from the target structure with high accuracy (low r.m.s.d.).

Only small part of actual model

Crystal of unknown structure

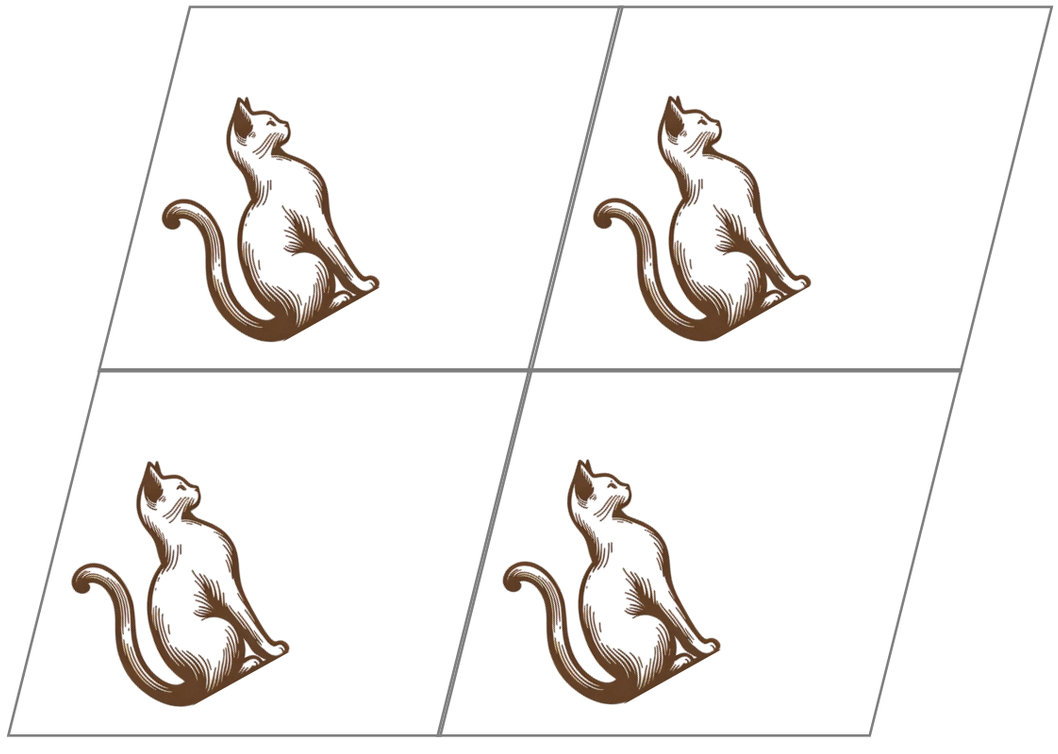


The search model

- Finding a suitable search model is critical step in MR.
- Should provide a high proportion of the scattering from the target structure with high accuracy (low r.m.s.d.).

Not similar to the target

Crystal of unknown structure



What can be used as search model?

Search model should be similar to the target structure.

1) **Homologue structure**
   If the sequence is similar → structures may be similar
   - Sequence comparison search
   - Prune regions of large sequence diversity
   - Truncate side-chains

Search model

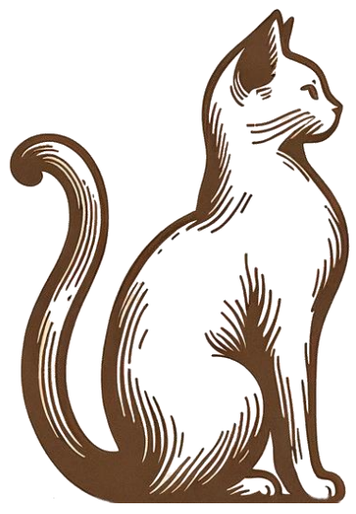
Unknown structure

What can be used as search model?

Search model should be similar to the target structure.

1) **Homologue structure**
   If the sequence is similar → structures may be similar
   - Sequence comparison search
   - Prune regions of large sequence diversity
   - Truncate side-chains

Search model



Unknown structure

What can be used as search model?

Search model should be similar to the target structure.

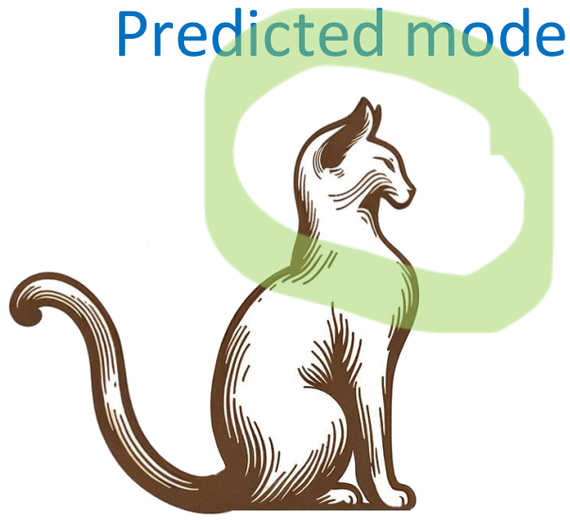
2) Predicted structure (AlphaFold, RosettaFold)

Process the predicted model before MR:

- Remove low pLDDT regions
- Split into domains if necessary
- Convert pLDDt to B-factors

Unknown structure

Predicted model



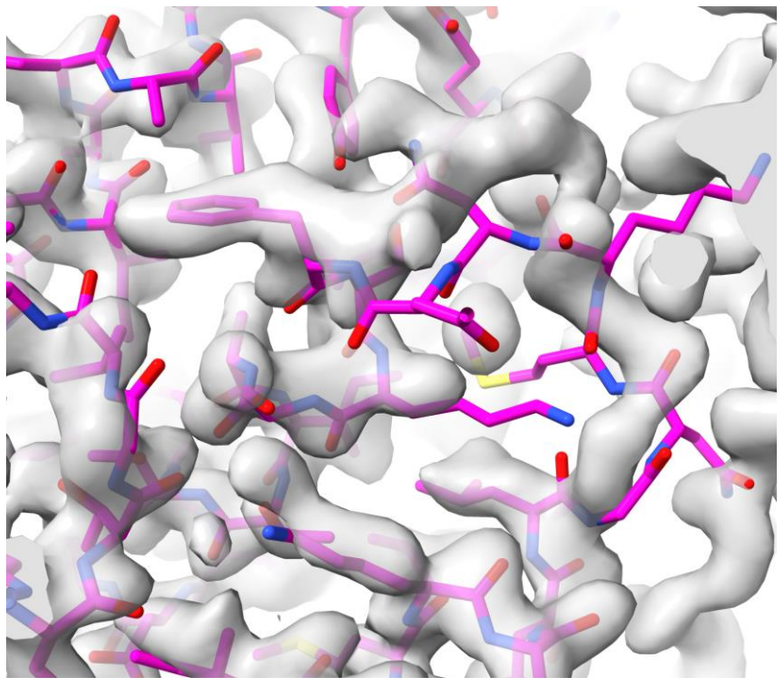
May have
- Distortions
- Incorrect domain relationships

Limitations of predicted models

<u>AI/ML model predictions</u>

- RNA/DNA not as accurate yet.
- High pLDDT regions are not necessarily correct.
- Know almost nothing about physics and chemistry.
- Generate a single structure, but no conformational changes (influenced by pH, temperature, ions, ligands, other proteins).

Wrong prediction



Molecular replacement: Scoring

Compare observed and calculated diffraction.

Poor score

Good score

Different approaches:

- Patterson function
- Maximum-likelihood Methods (Phaser)

MR Scoring: Maximum Likelihood Method

"For any postulated orientation and position of the model, what is the probability of obtaining the structure amplitudes that we observe?"

Explicitly models errors
- Experimental uncertainties
- r.m.s. coordinate error of the search model

→ Likelihood methods are more robust and generally give clearer solutions in difficult cases

Maximum Likelihood Scoring in Phaser

LLG = Log Likelihood gain

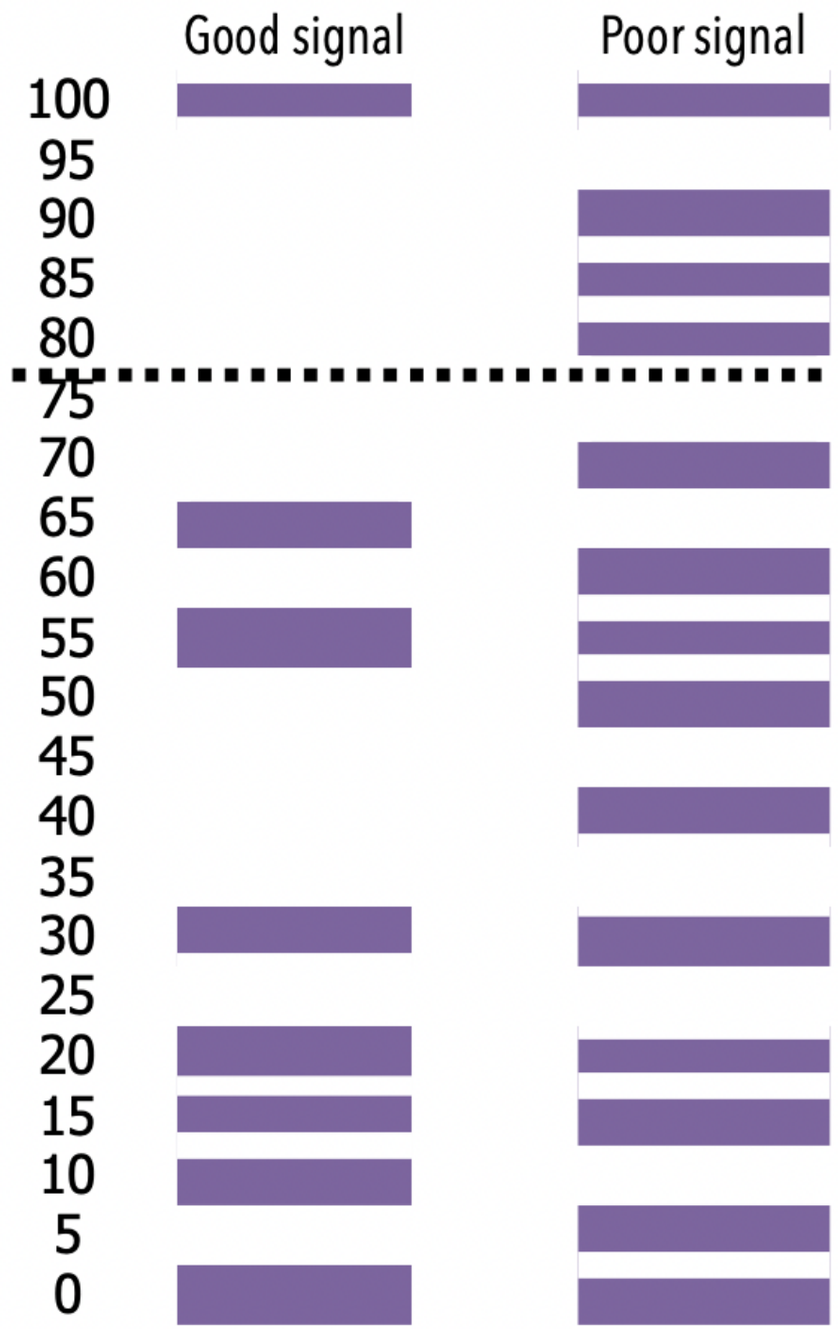
→ It measures how much better the data can be predicted with the search model than with a random distribution of the same atoms.

TF-Z = how many standard deviations your solution is above the mean (the higher the better).

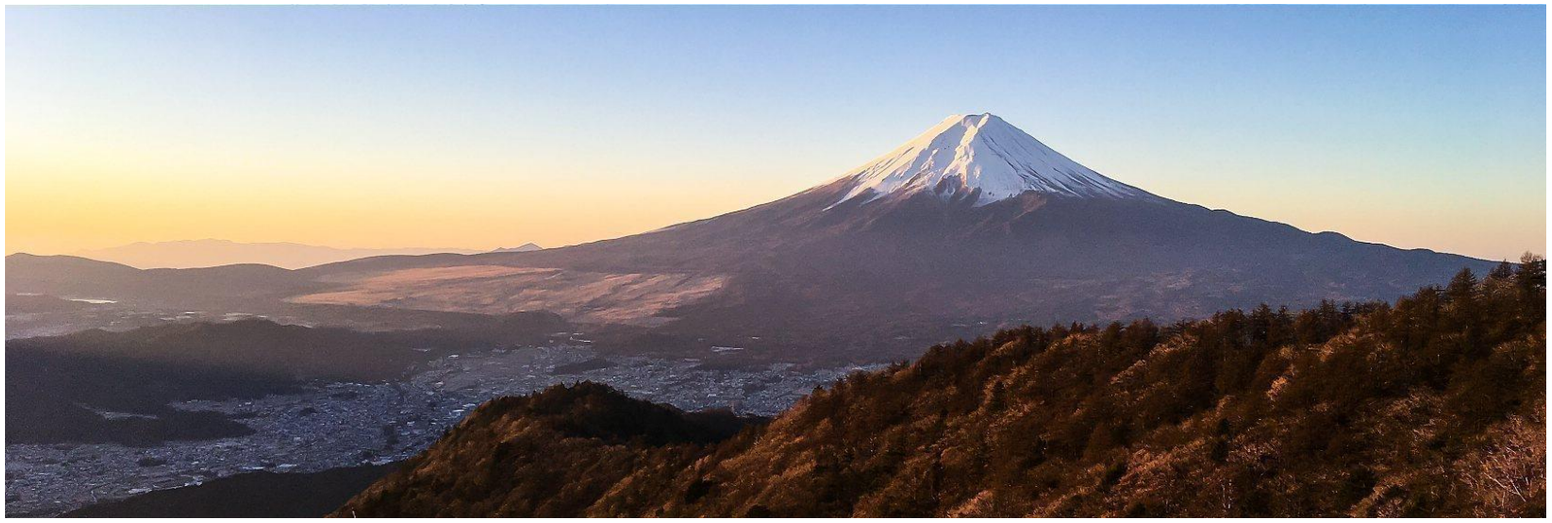
Maximum Likelihood Scoring in Phaser

Select solutions that are over 75% of the difference between the top peak and the mean.

- Good signal, few potential solutions
- Poor signal, many potential solutions

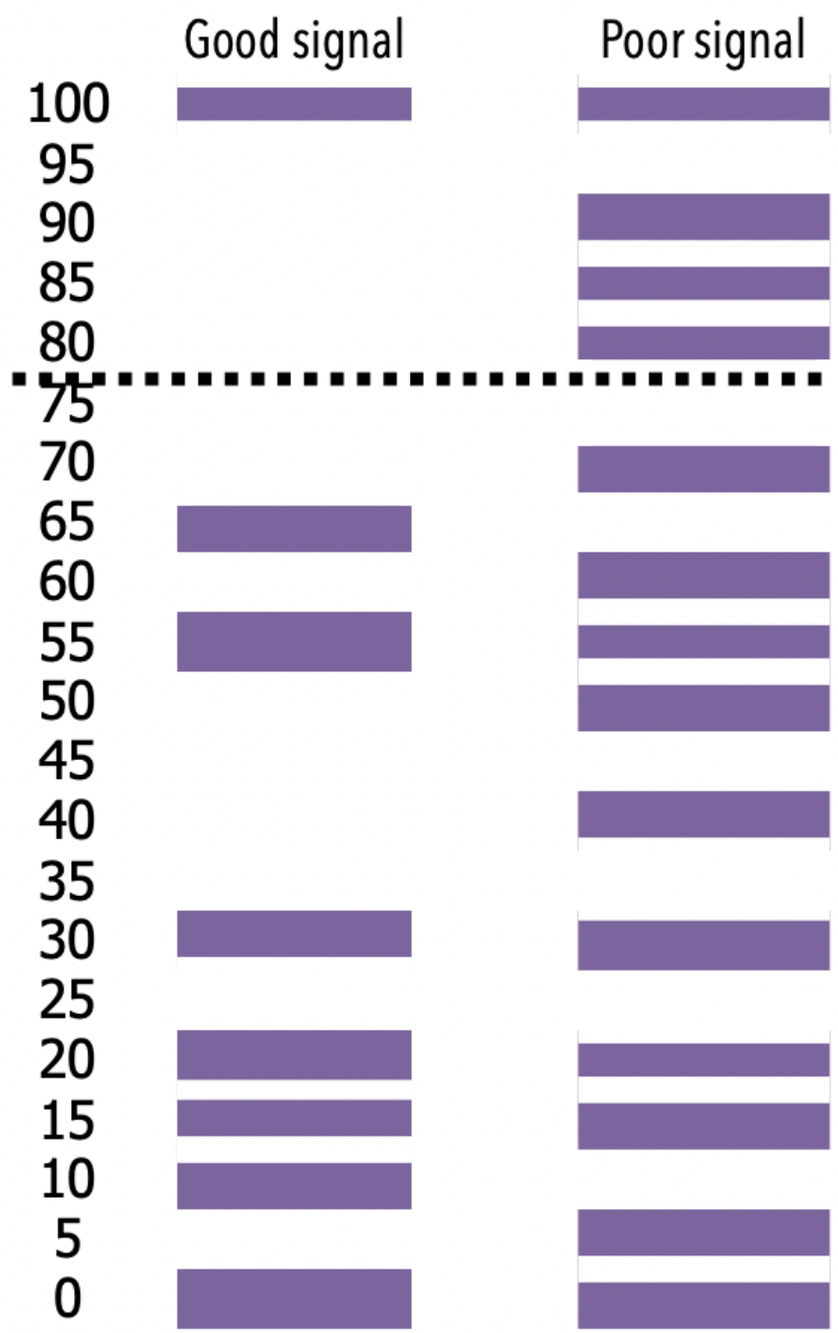


Maximum Likelihood Scoring in Phaser



Good signal

Poor signal



Maximum Likelihood Scoring in Phaser

TF Z-score	LLG score	Solved?
< 5	< 25	no
5 - 6	25 - 36	unlikely
6 - 7		
7 - 8		
> 8		

Summary

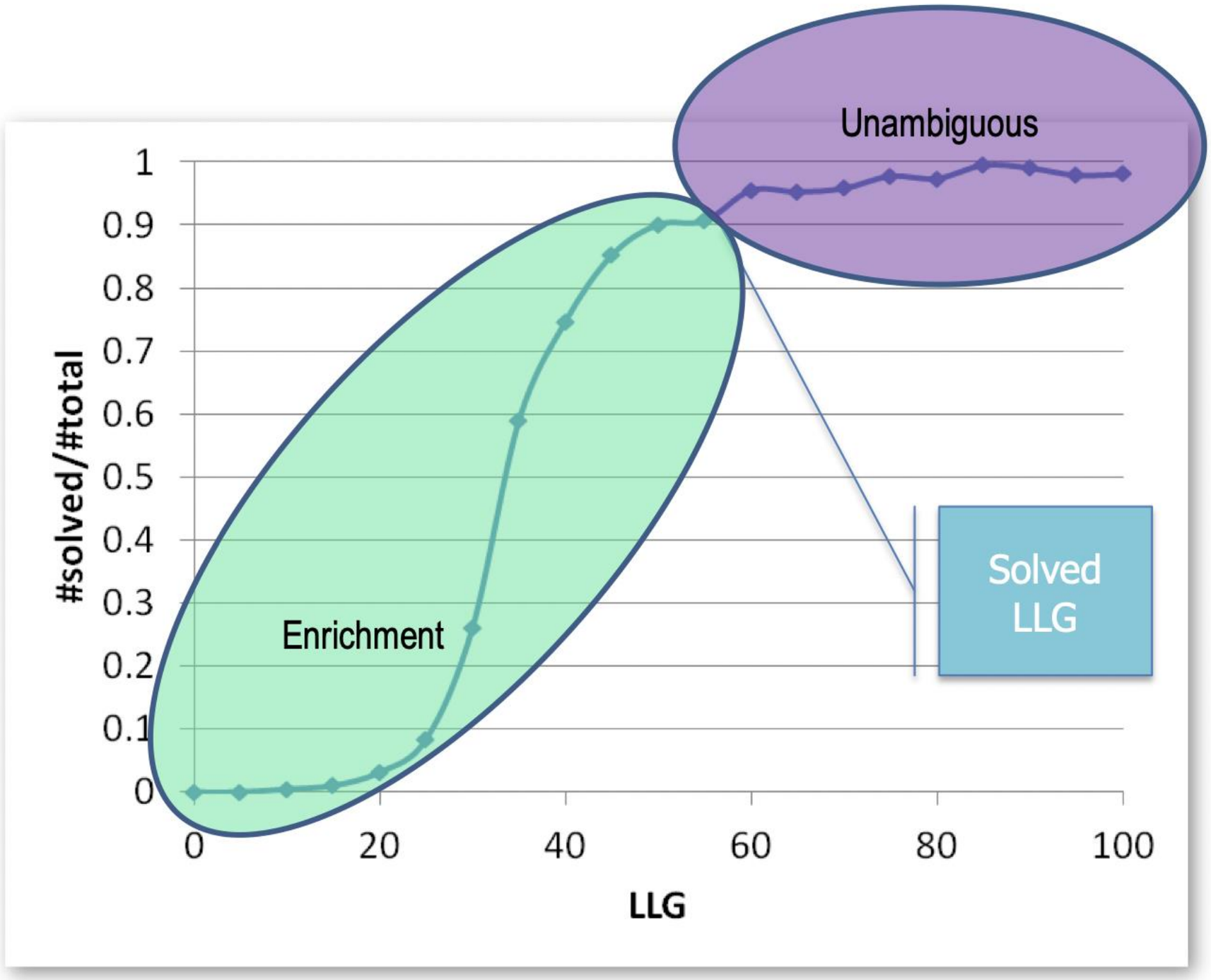
Phaser has found 1 MR solution(s).

Top LLG: 11480.402

Top TFZ: 68.1

Spacegroup: C 1 2 1

Maximum Likelihood Scoring in Phaser



Plot of LLG versus success in structure solution

Database of over 23000 MR problems

R.D. Oeffner

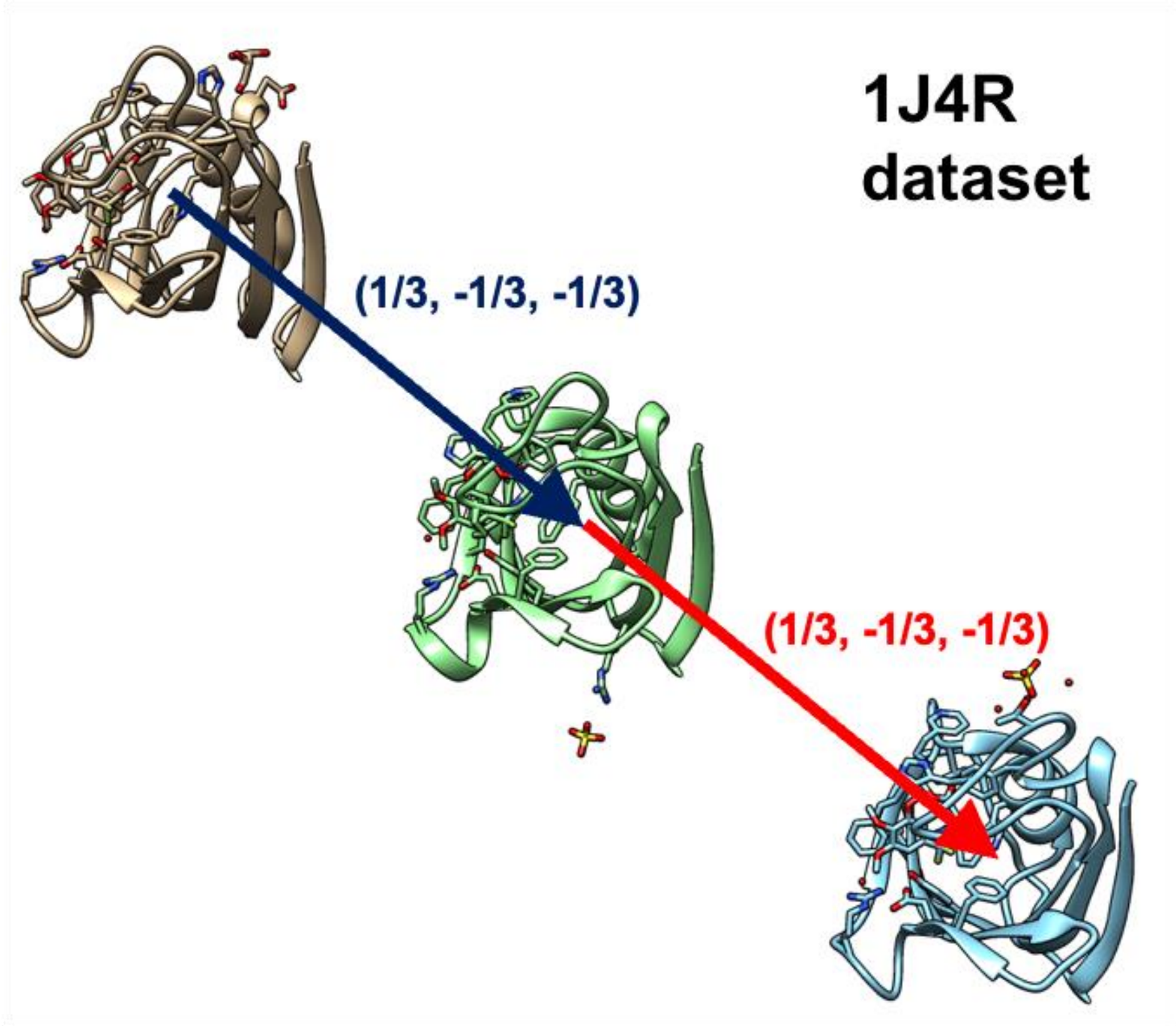
Likelihood is sensitive...

- ...to correct orientation and position of molecular replacement model
  - successful in solving structures with distant relatives, small fragments, or many copies in asymmetric unit
- ...to violations of assumptions
  - data implicitly assumed to be isotropic
    → important to account for anisotropy
  - tNCS: intensity distribution is modulated
  - components may not be equally well-ordered
    →important to correct for differences in overall B-factors

Slide from Randy Read

Translational NCS

Two or more copies of the molecule are related by pure translation.

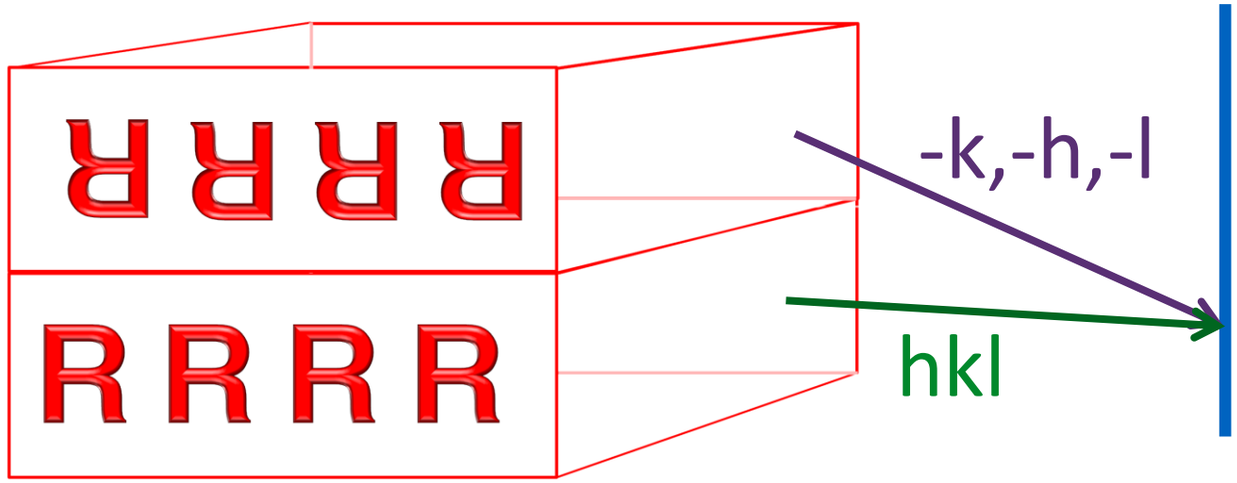


Strong peak in Patterson function

In this case at (1/3, -1/3, -1/3)

Twinning

Identical but rotated crystals sandwiched together



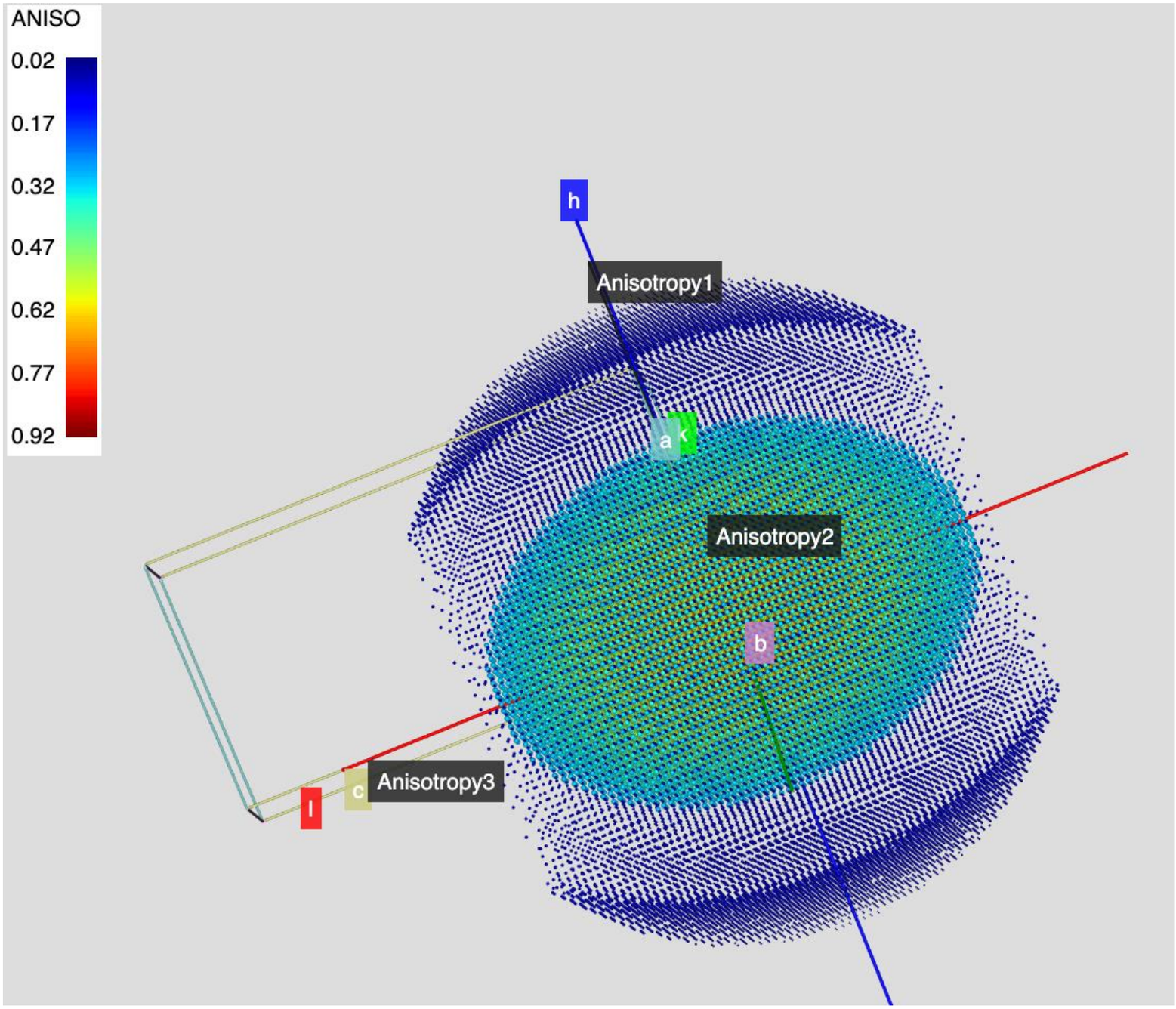
…is mixed with twin-law related reflection (e.g., -k,-h,-l)

Diffraction spot for (h,k,l) reflection…

Possible twin laws depend on your crystal symmetry and cell dimensions

Anisotropic data

Crystal does not diffract equally well in all directions.



Phaser takes care of anisotropy corrections → No need to truncate (Truncating may confuse Phaser's intensity analyses)

What is needed to run Phaser MR

- Reflection data
- Search model
- Error estimation of the search model
  - Homologue: sequence identity
  - Predicted model: r.m.s.d. (1.0 Å)
- ASU content:
  - Sequence of the construct
  - How many copies of the molecule(s)
- Awareness of data pathologies:
  - Twinning
  - tNCS
  - anisotropy

If MR does not work

- Double check space group (twinning).
- Does your data set have uncurable pathologies? (low resolution, high mosaicity)
- Is the content correct (e.g., impurities)?

  Is it the correct sequence file?

- Can the search model be improved?
  - is the search model complete enough (AF multimer)?
  - use other prediction software.
  - use a different splitting method.
- Use secondary structure based software
  - ARCIMBOLDO, AMPLE
- Discuss with your colleagues; ask at bulletin boards (ccp4bb); ask software developers

The Phenix Project

Lawrence Berkeley Laboratory

Paul Adams, Pavel Afonine, Dorothee Liebschner, Nigel Moriarty, Billy Poon, Christopher Schlicksup, Oleg Sobolev

Los Alamos National Laboratory
New Mexico Consortium

Tom Terwilliger, Li-Wei Hung

UTHealth

Matt Baker

University of Cambridge

Randy Read, Airlie McCoy, Alisia Fadini

Duke University

Jane & David Richardson, Christopher Williams, Vincent Chen

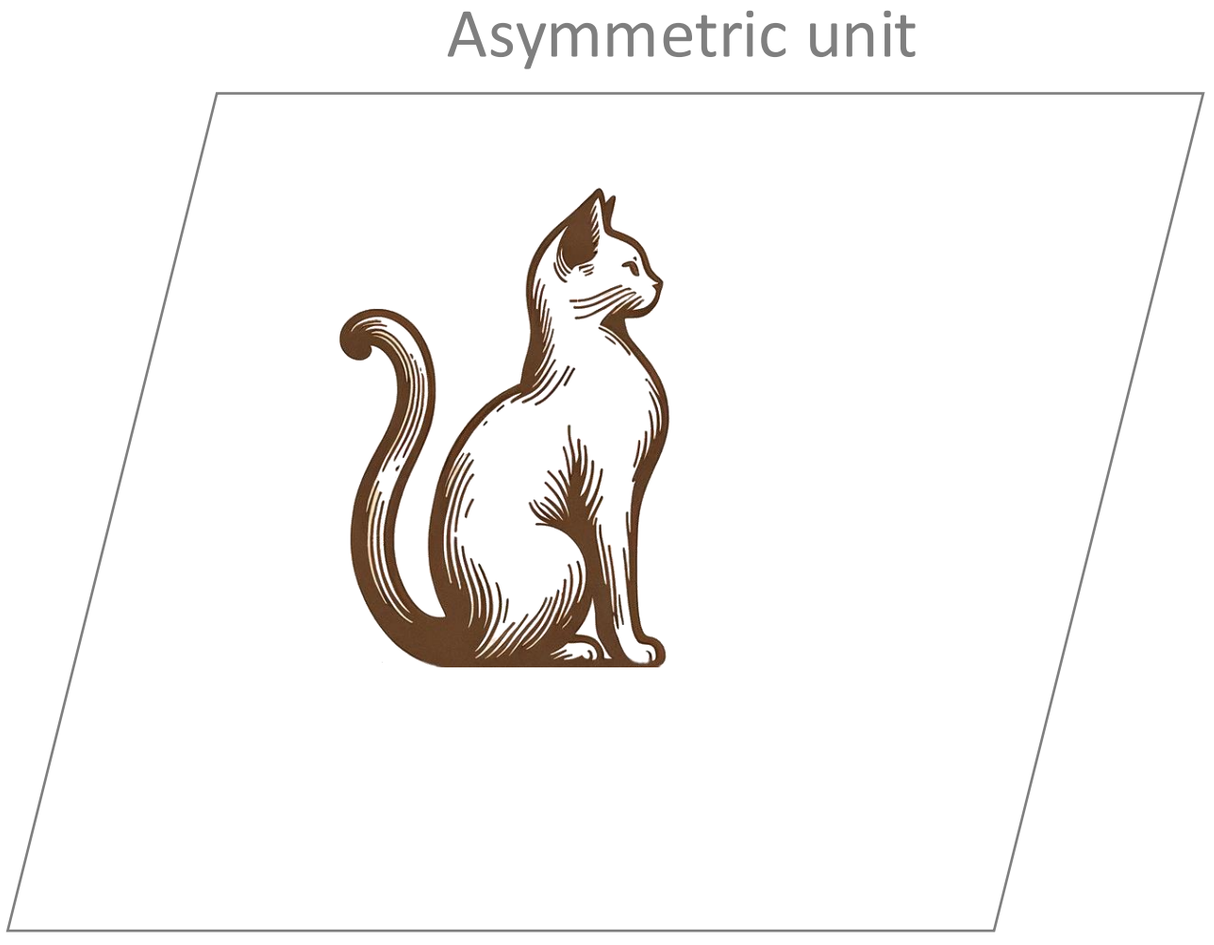
An NIH/NIGMS funded Program Project

Liebschner D, *et al*., Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. Acta Cryst. 2019 **D75**:861–877

ASU content

- Sequence of your construct
- How many copies of the the molecule(s)?

N = 1

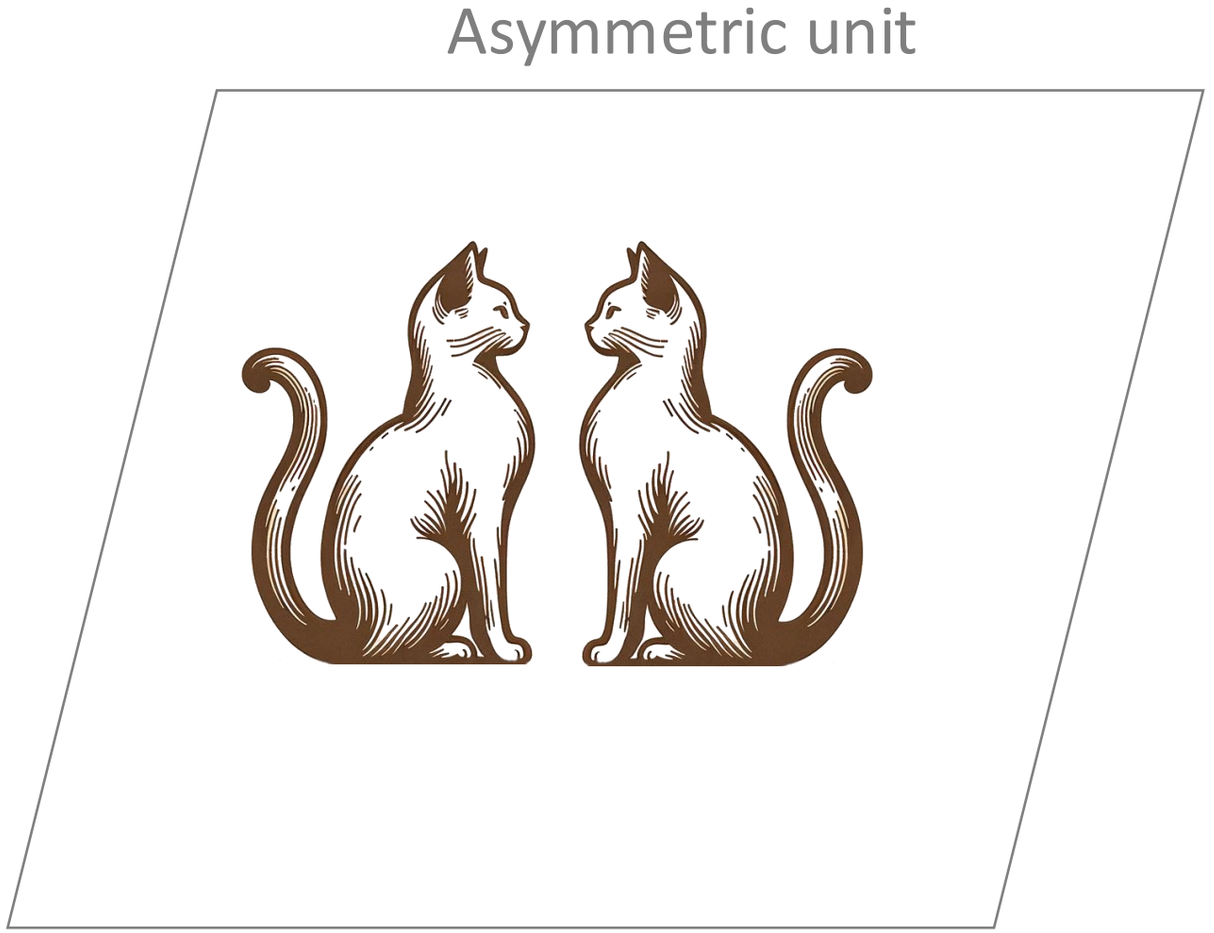


One copy in the asymmetric unit.

ASU content

- Sequence of your construct
- How many copies of the the molecule(s)?

N = 2

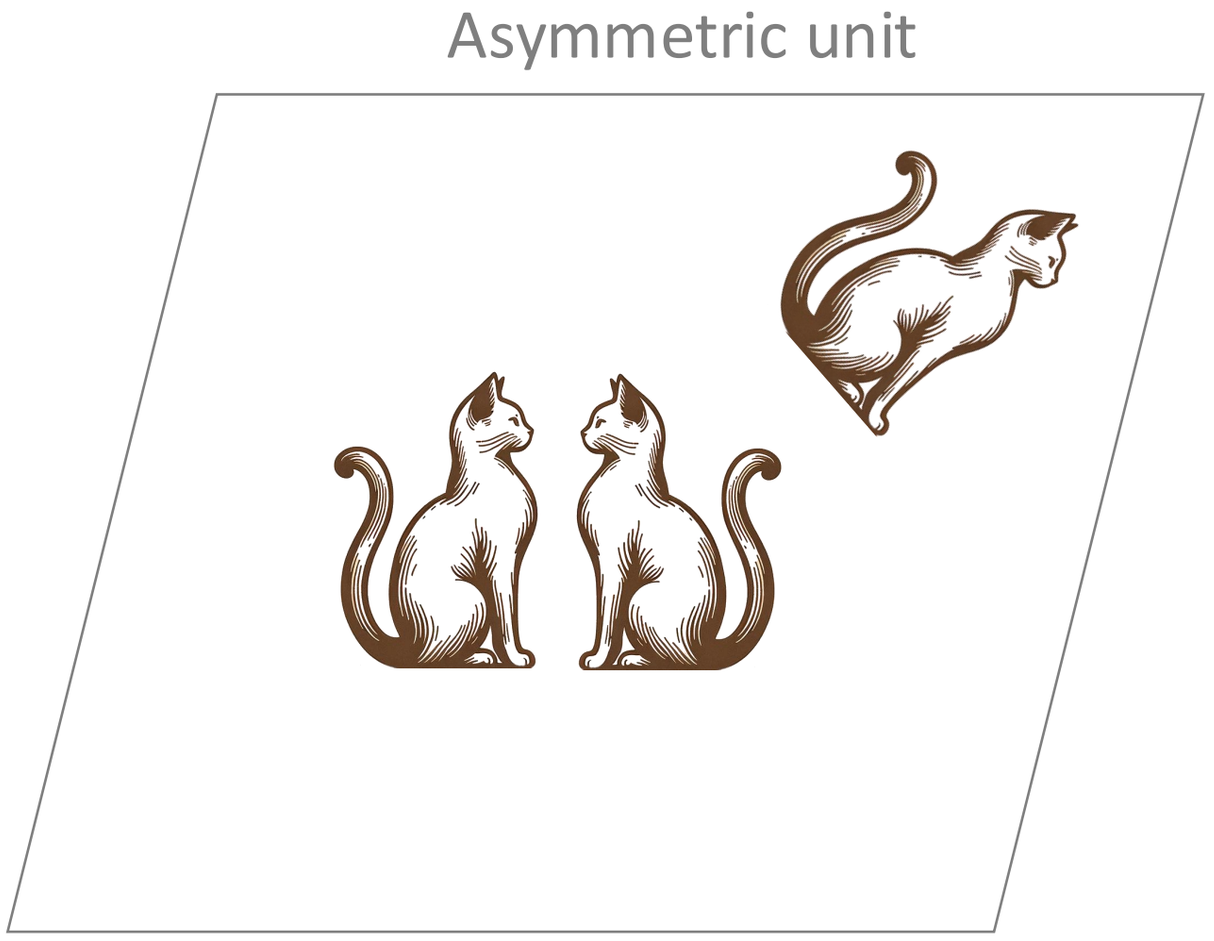


Two copies in the asymmetric unit, rotational symmetry.

ASU content

- Sequence of your construct
- How many copies of the the molecule(s)?

N = 3

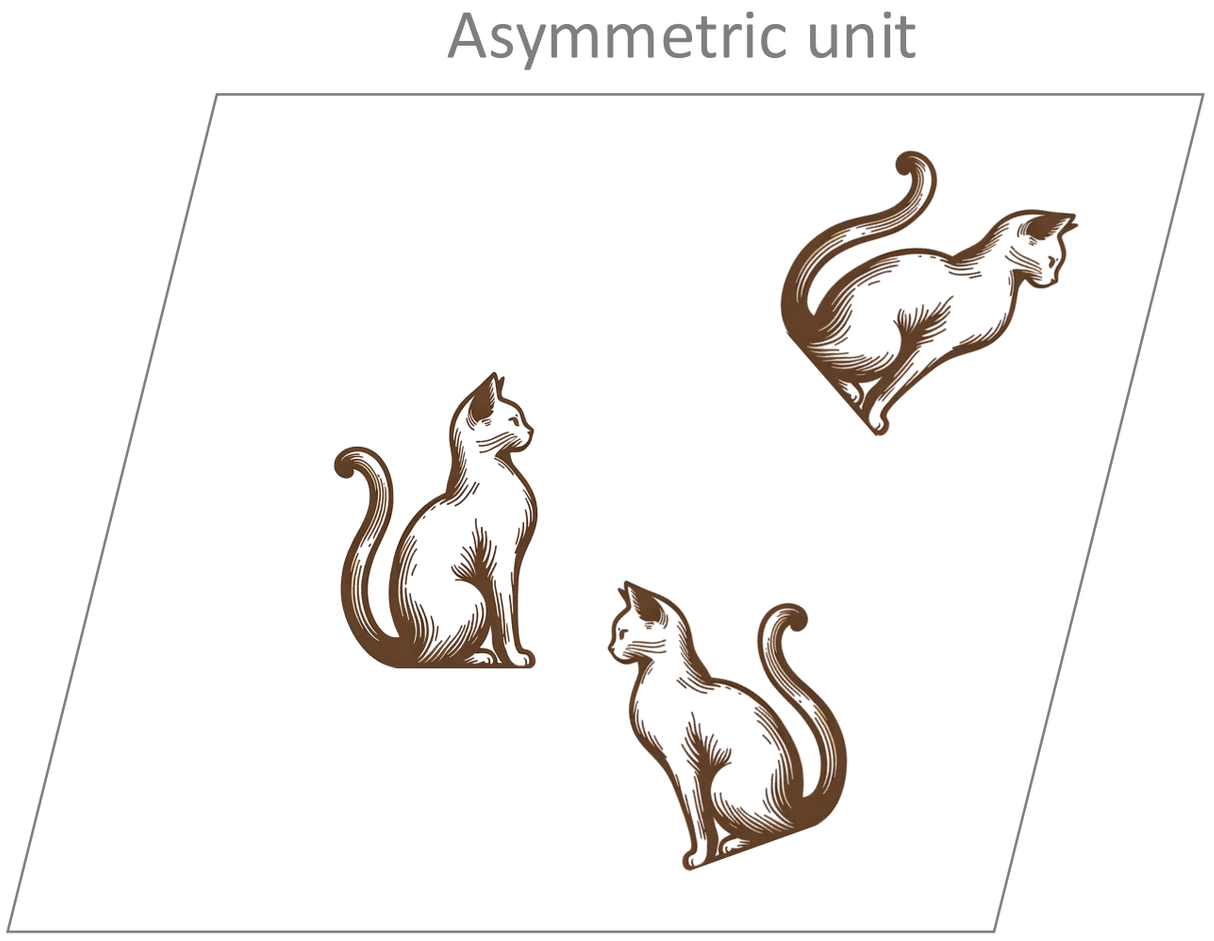


Three copies in the asymmetric unit, two in rotational symmetry, one unrelated.

ASU content

- Sequence of your construct
- How many copies of the the molecule(s)?

N = 3

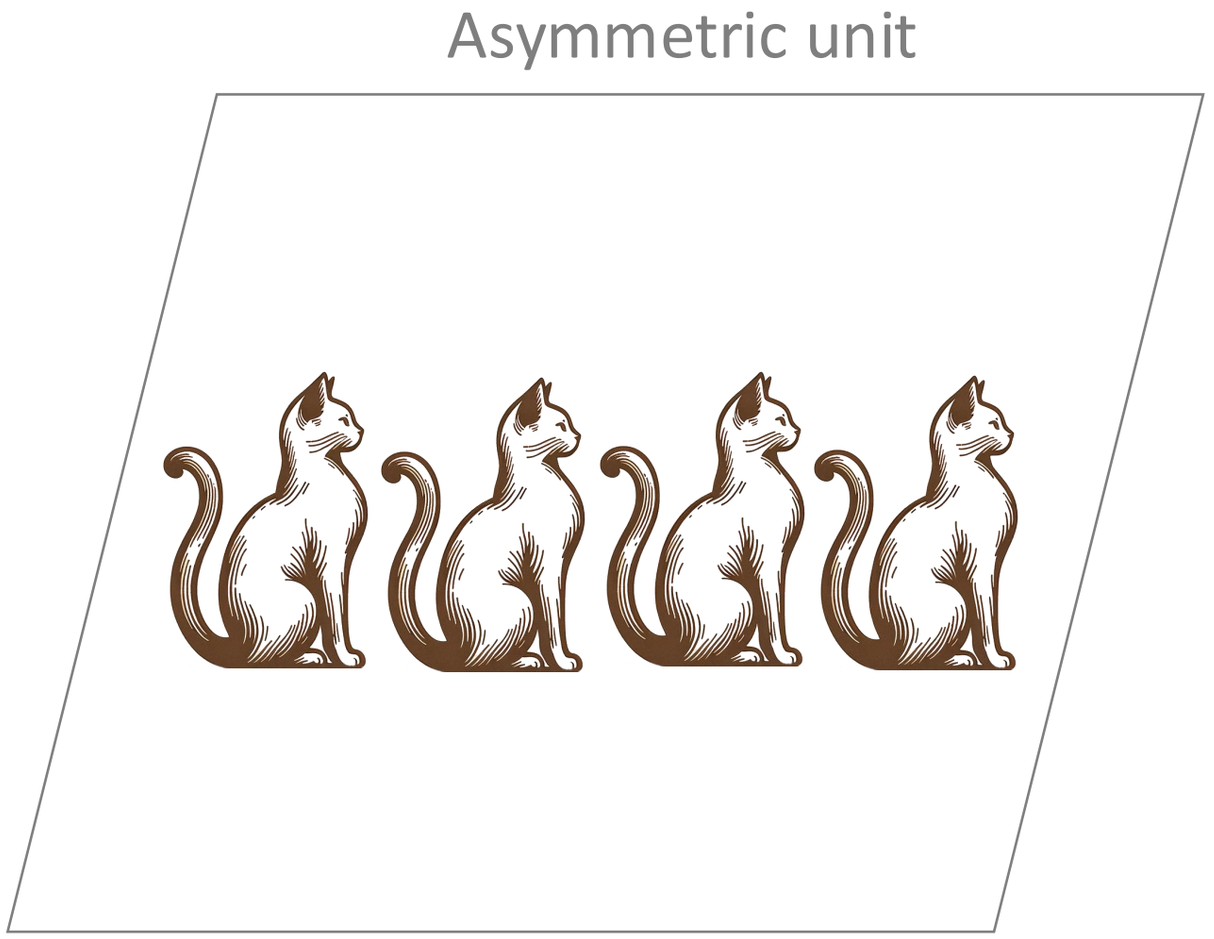


Three copies in the asymmetric unit.

ASU content

- Sequence of your construct
- How many copies of the the molecule(s)?

N = 4

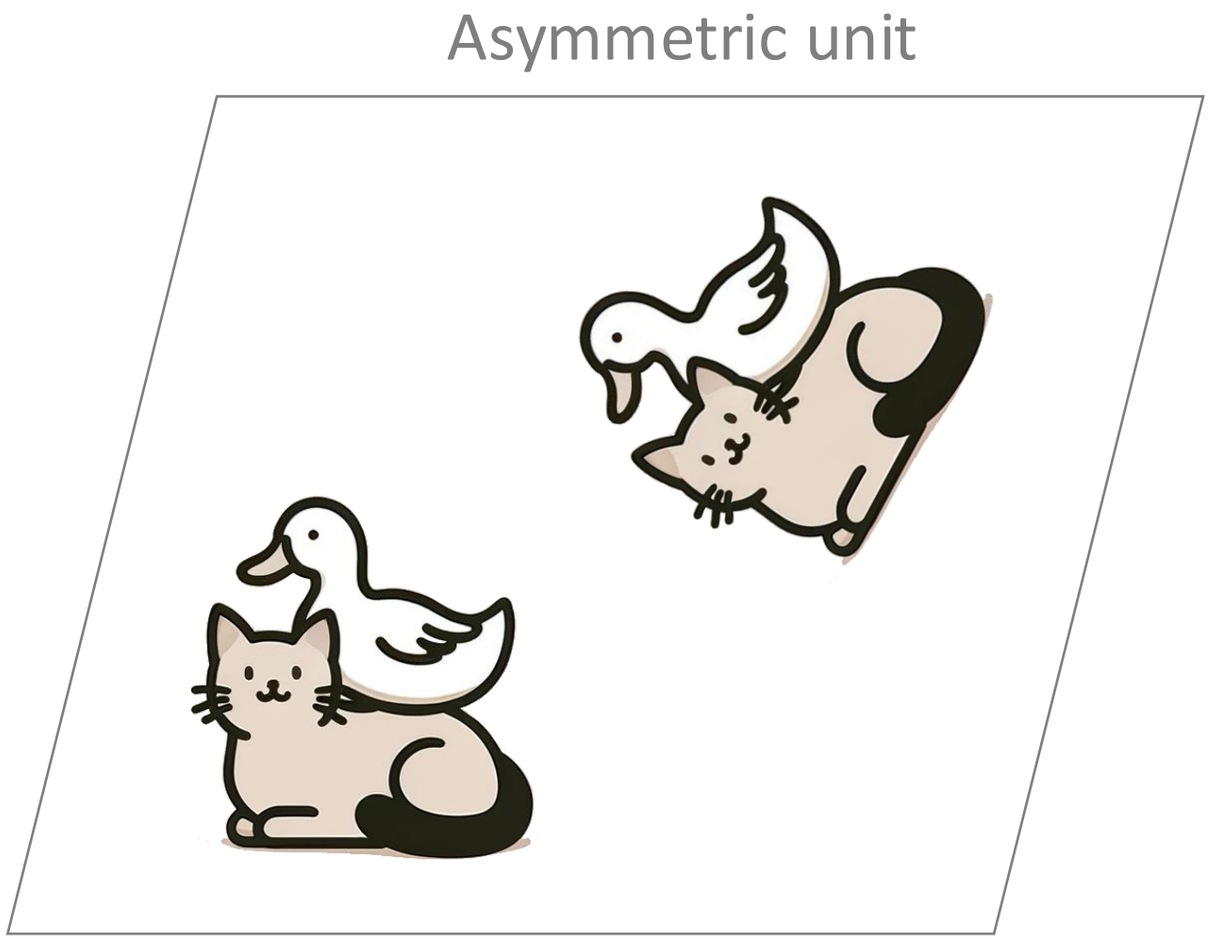


Four copies in the asymmetric unit, translational non-crystallographic symmetry.

ASU content

- Sequence of your construct
- How many copies of the the molecule(s)?
- Different domains?

N = 2



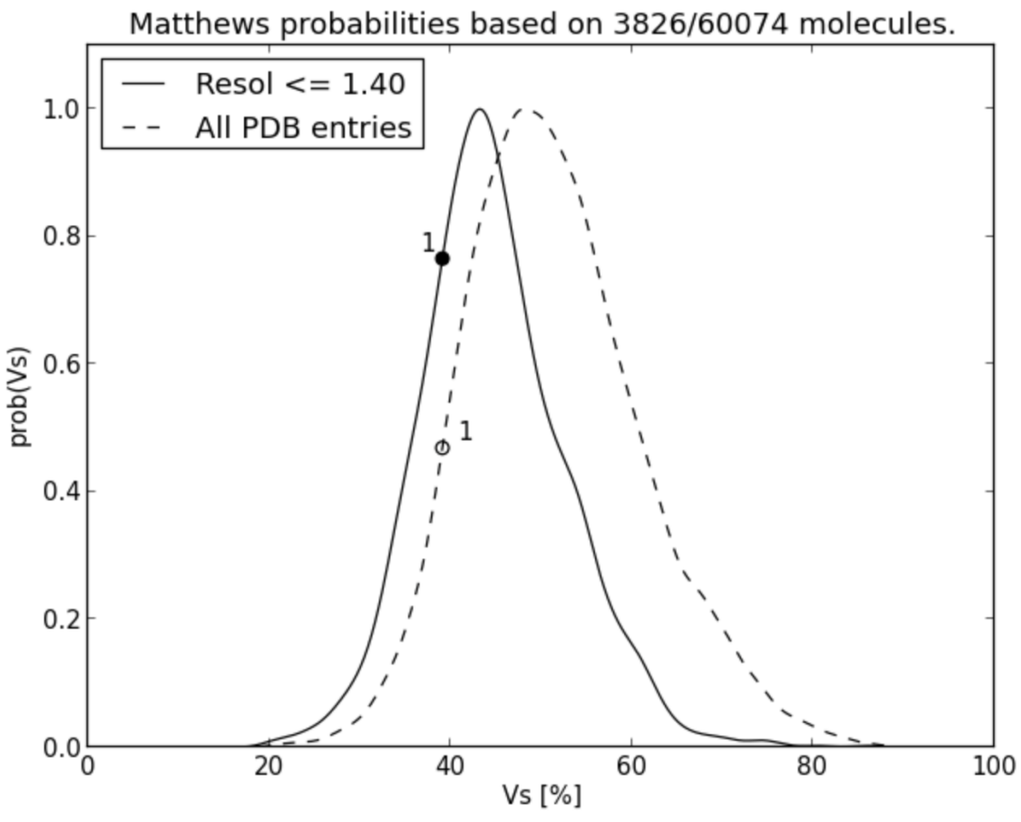
If mutual orientation is unknown, search for the domains separately.

Estimating the number of molecules in the ASU

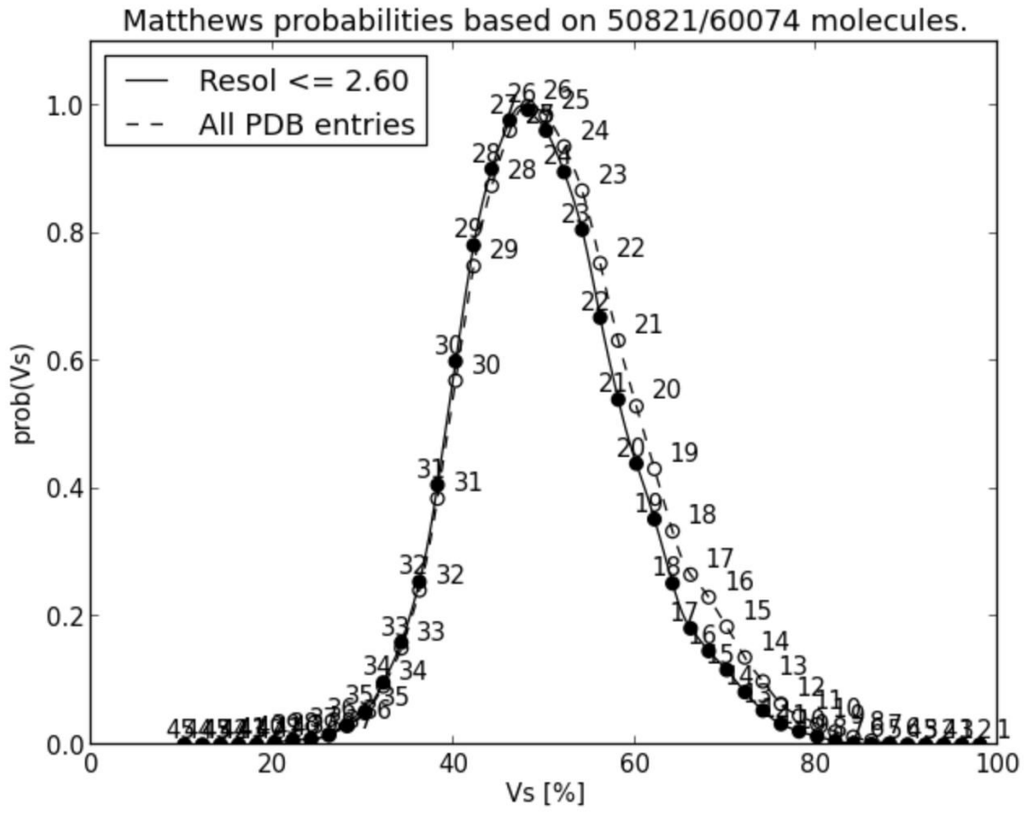
Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

Compare your value of V_M with histograms for known structures (from PDB) → choose most probable.

Few possibilities

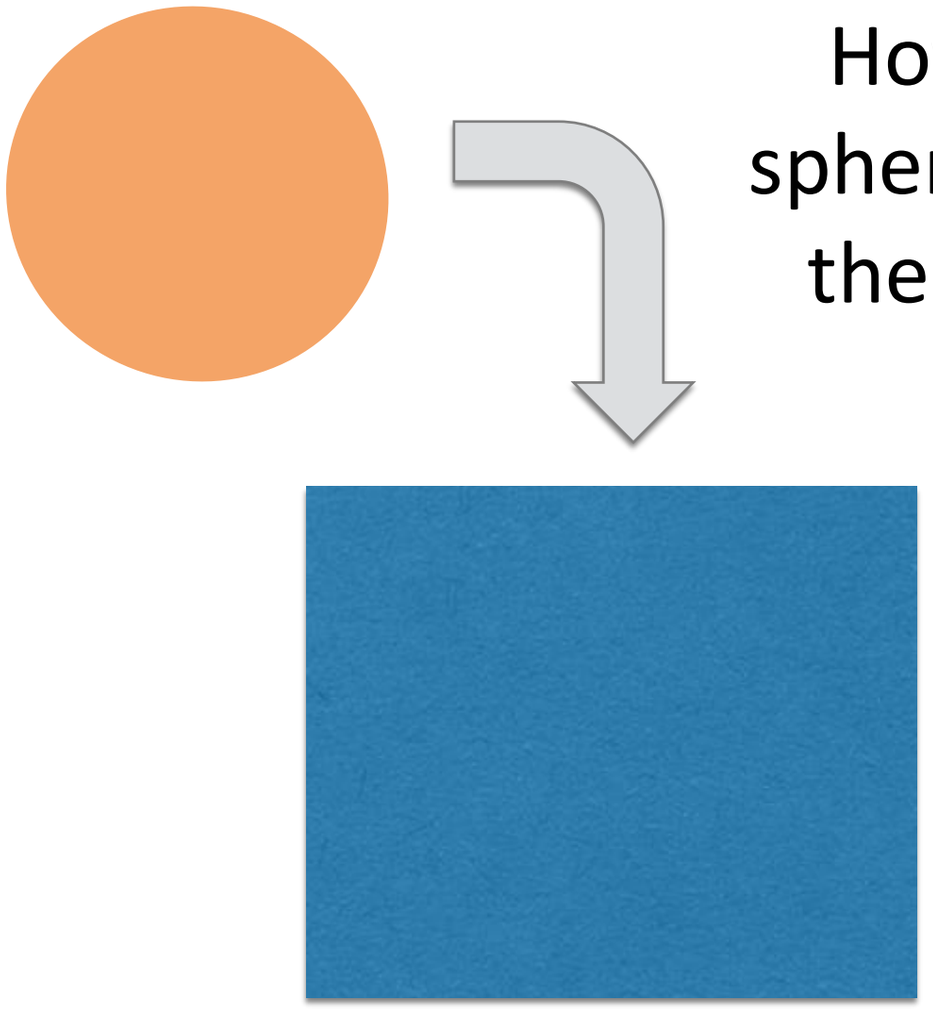


Many possibilities



Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

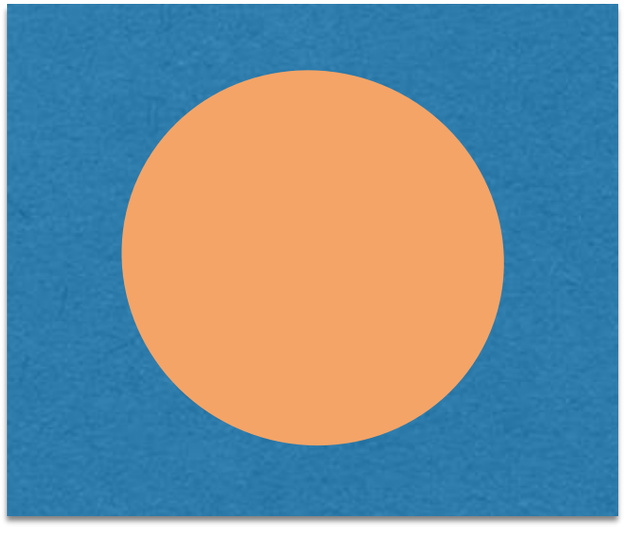


How many spheres fit into the square?

Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

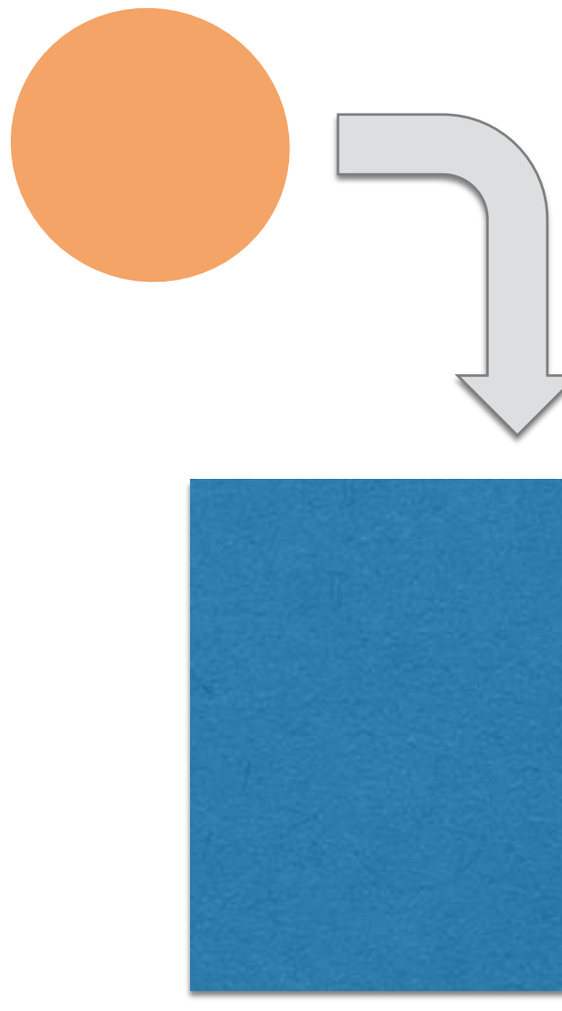
How many spheres fit into the square?



Clearly only one sphere

Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

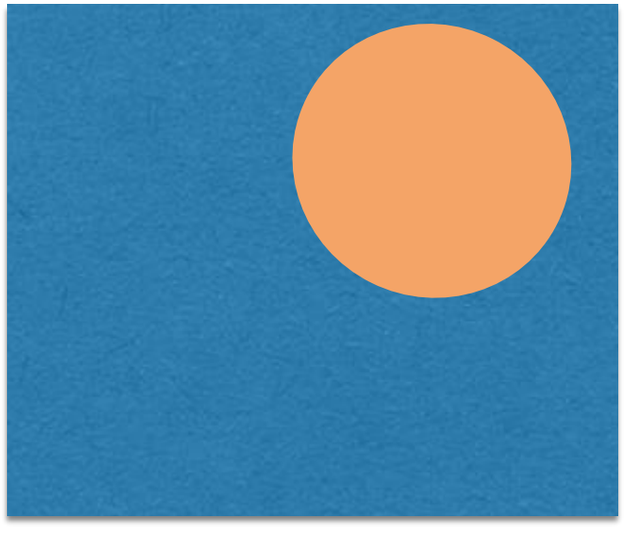


How many spheres fit into the square?

Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

How many spheres fit into the square?

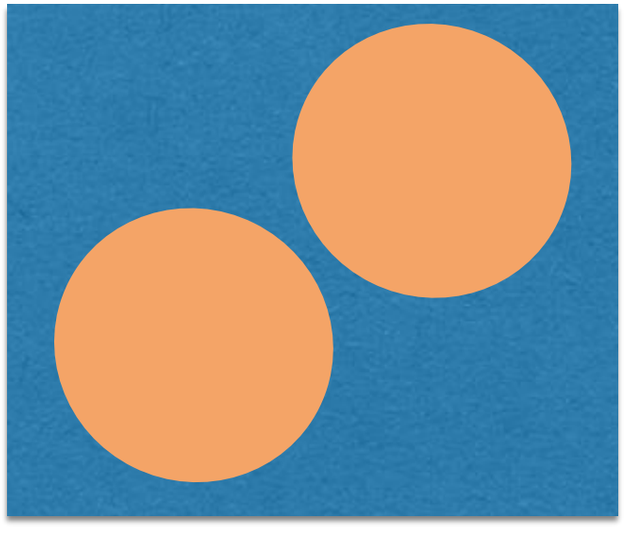


There could be only one…

Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

How many spheres fit into the square?

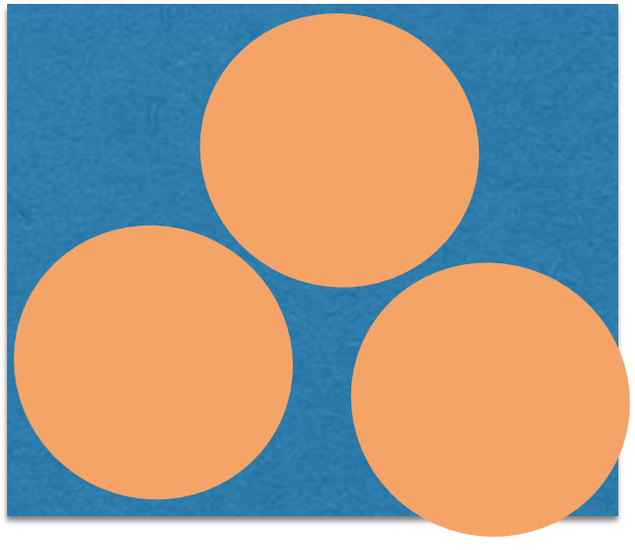


… or maybe two

Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

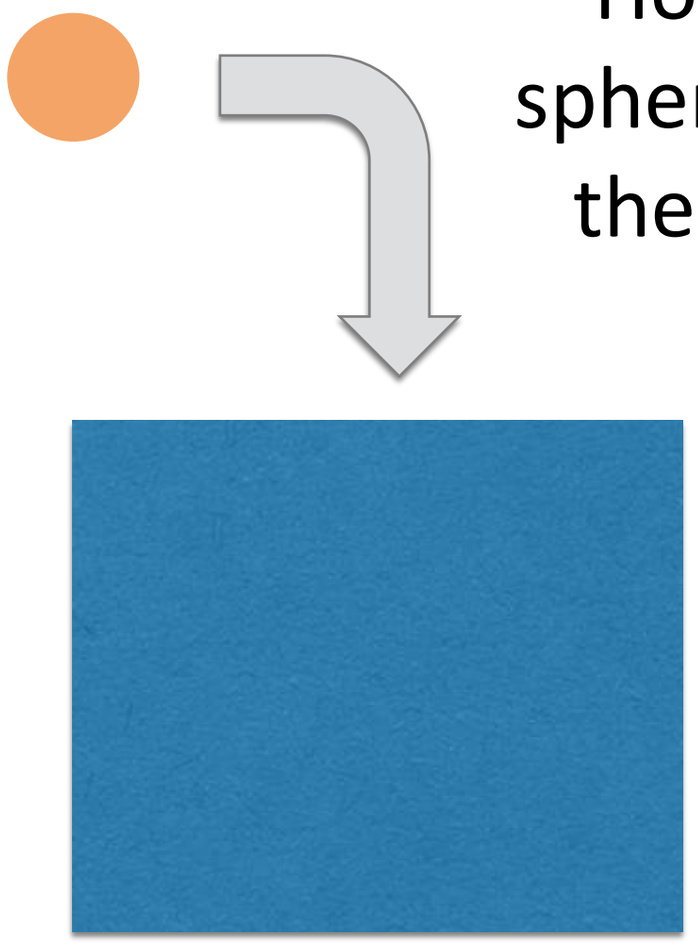
How many spheres fit into the square?



…but not three

Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

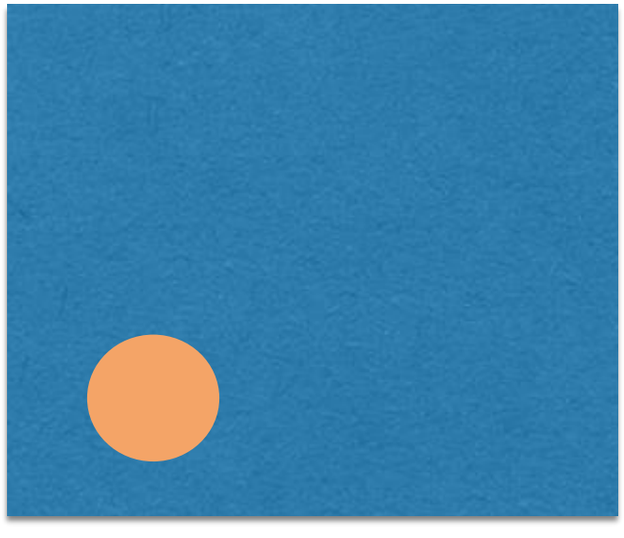


How many spheres fit into the square?

Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

How many spheres fit into the square?

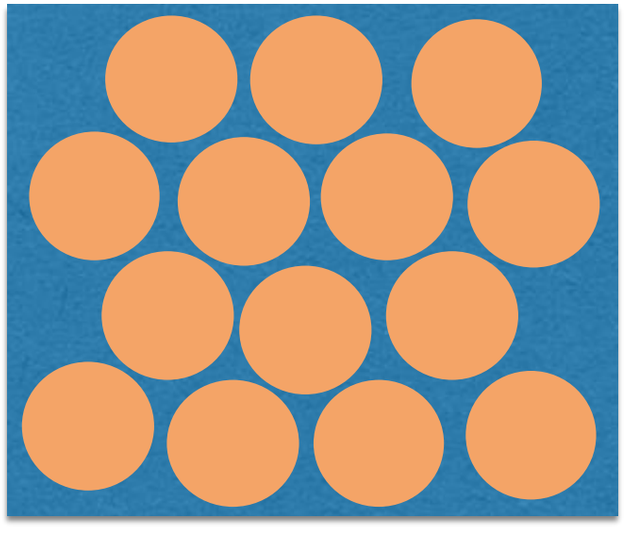


Only one

Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

How many spheres fit into the square?

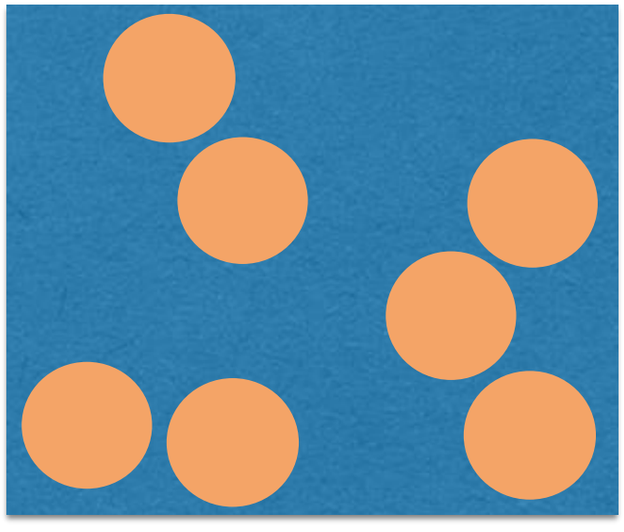


Many (14)

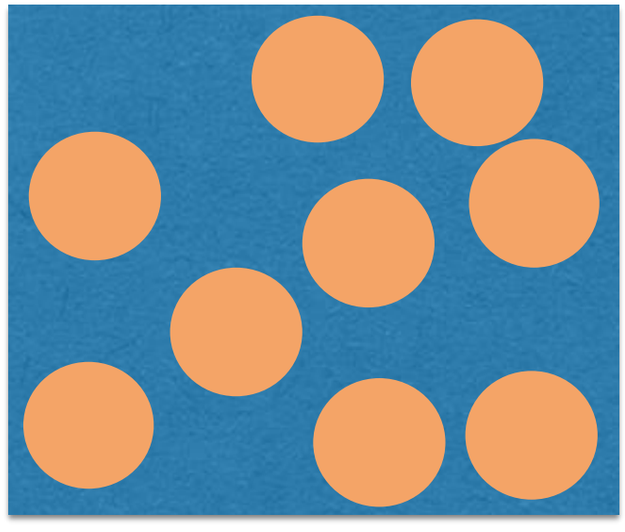
Estimating the number of molecules in the ASU

Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

How many spheres fit into the square?



7



9

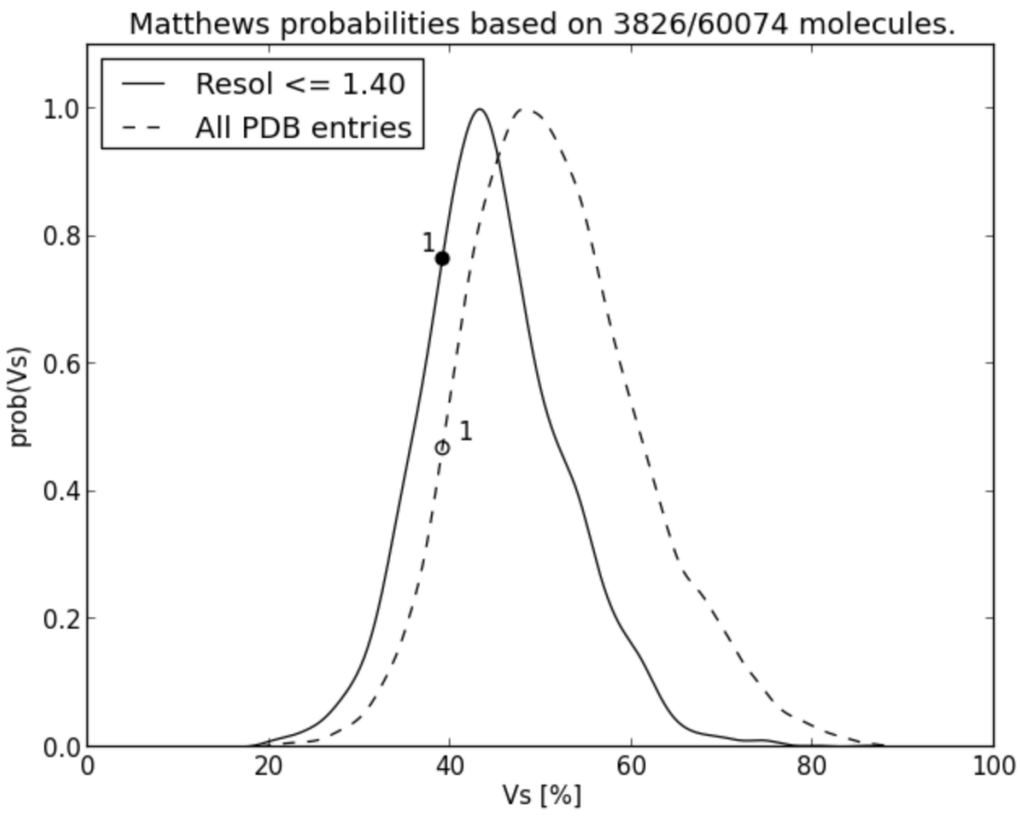
Most likely: something in between

Estimating the number of molecules in the ASU

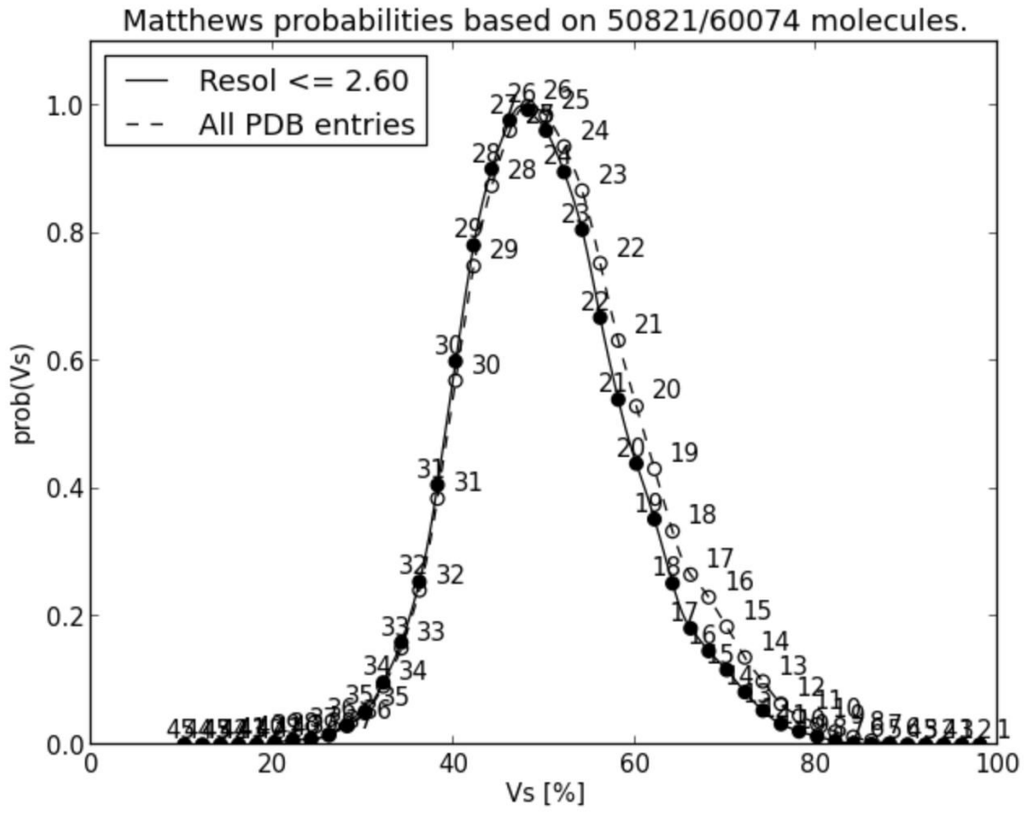
Matthews coefficient: $$V_M = \frac{\text{volume of asymmetric unit}}{\text{molecular weight}}$$

Compare your value of V_M with histograms for known structures (from PDB) → choose most probable.

Few possibilities



Many possibilities



MR with an AlphaFold model

POMGNT2 = protein O-linked mannose acetylglucosaminyl transferase-2

1) check Matthews coefficient

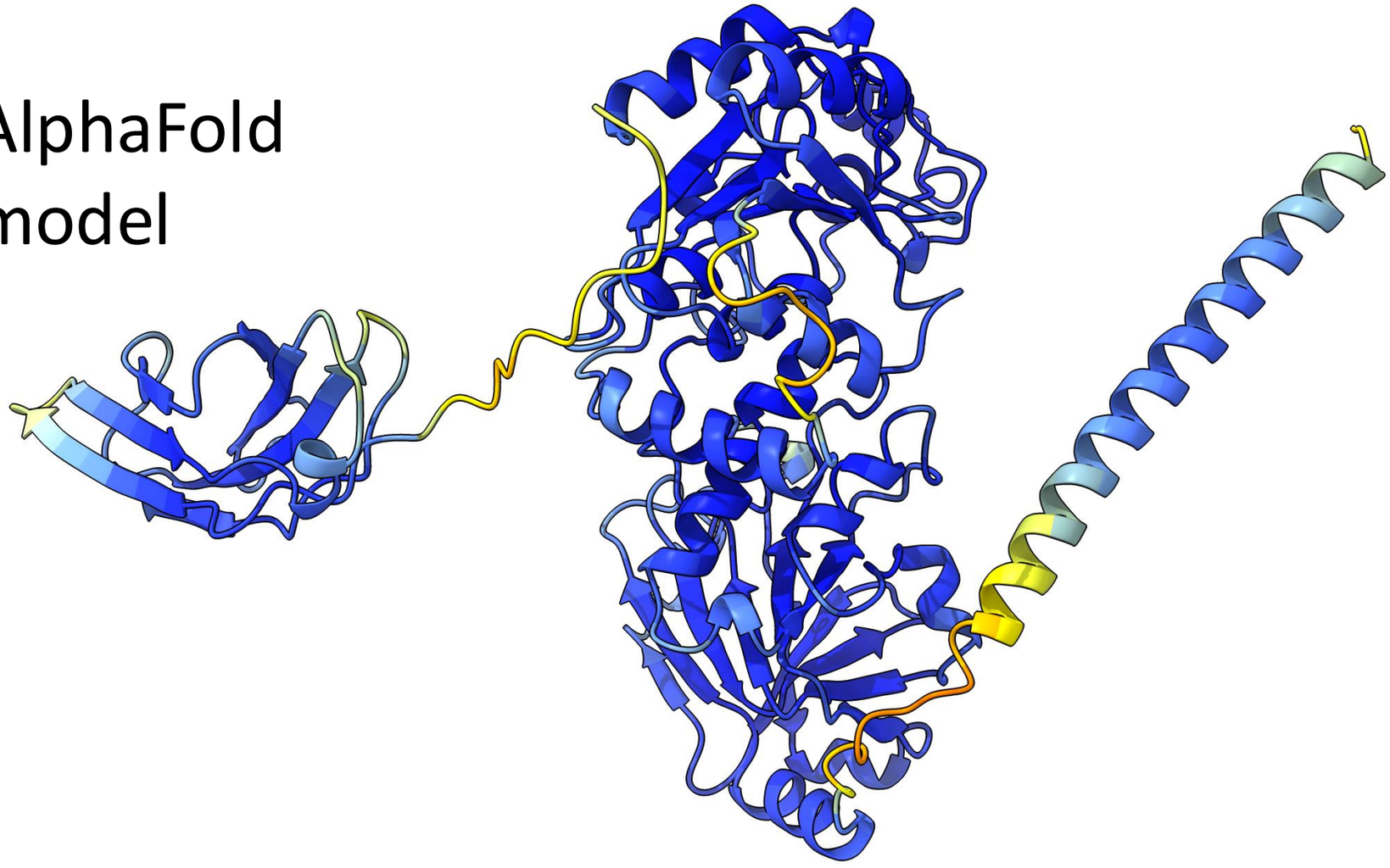
2) process predicted model

3) run Phaser molecular replacement

MR with an AlphaFold model

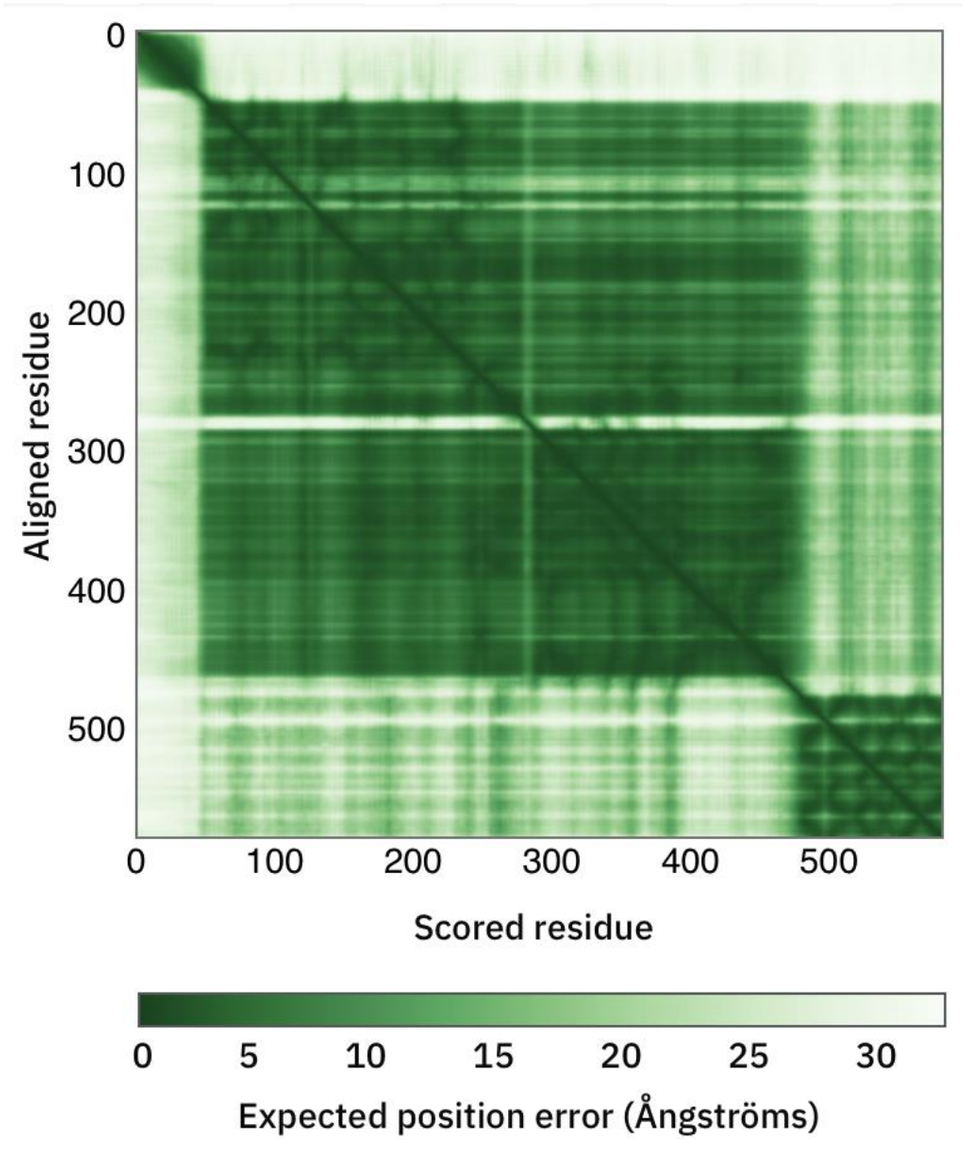
POMGNT2 = protein O-linked mannose acetylglucosaminyl transferase-2

AlphaFold model



MR with an AlphaFold model

POMGNT2 = protein O-linked mannose acetylglucosaminyl transferase-2



Predicted alignment error (PAE) plot

→ Dark green: the predicted relationship between a pair of residues is likely to be accurate

→ Suggests 3 domains

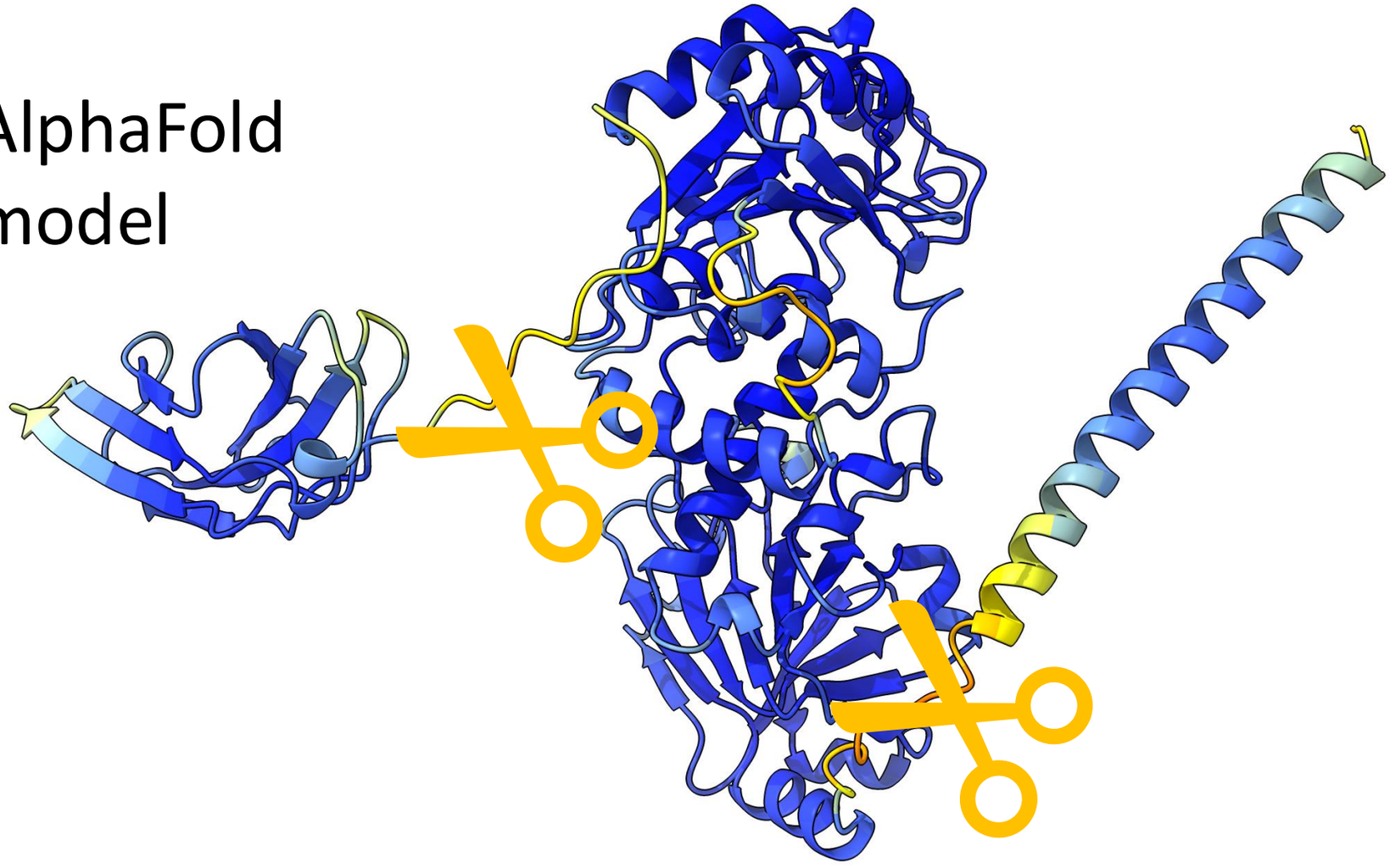
→ Mutual configuration is not so clear

MR with an AlphaFold model

POMGNT2 = protein O-linked mannose acetylglucosaminyl transferase-2

→ Remove low pLDDT regions, split into domains

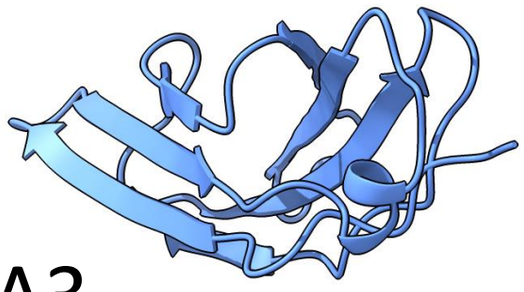
AlphaFold model



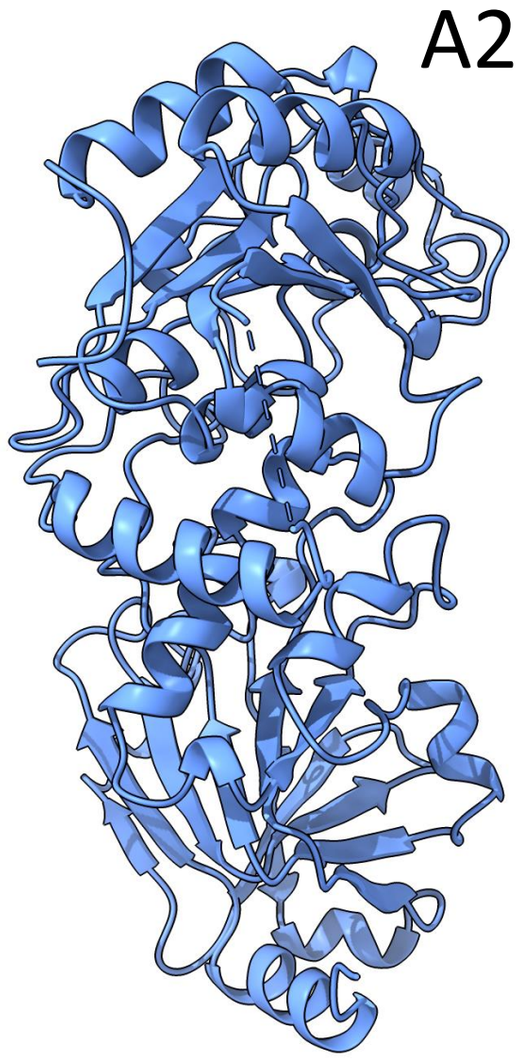
MR with an AlphaFold model

POMGNT2 = protein O-linked mannose acetylglucosaminyl transferase-2

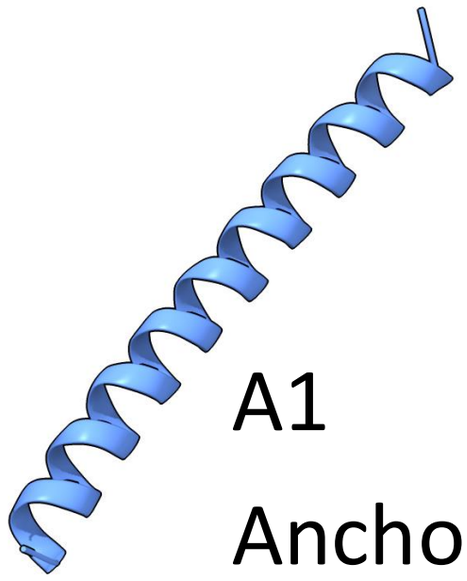
Processed AlphaFold model



A3



A2



A1

Anchor helix not in the construct